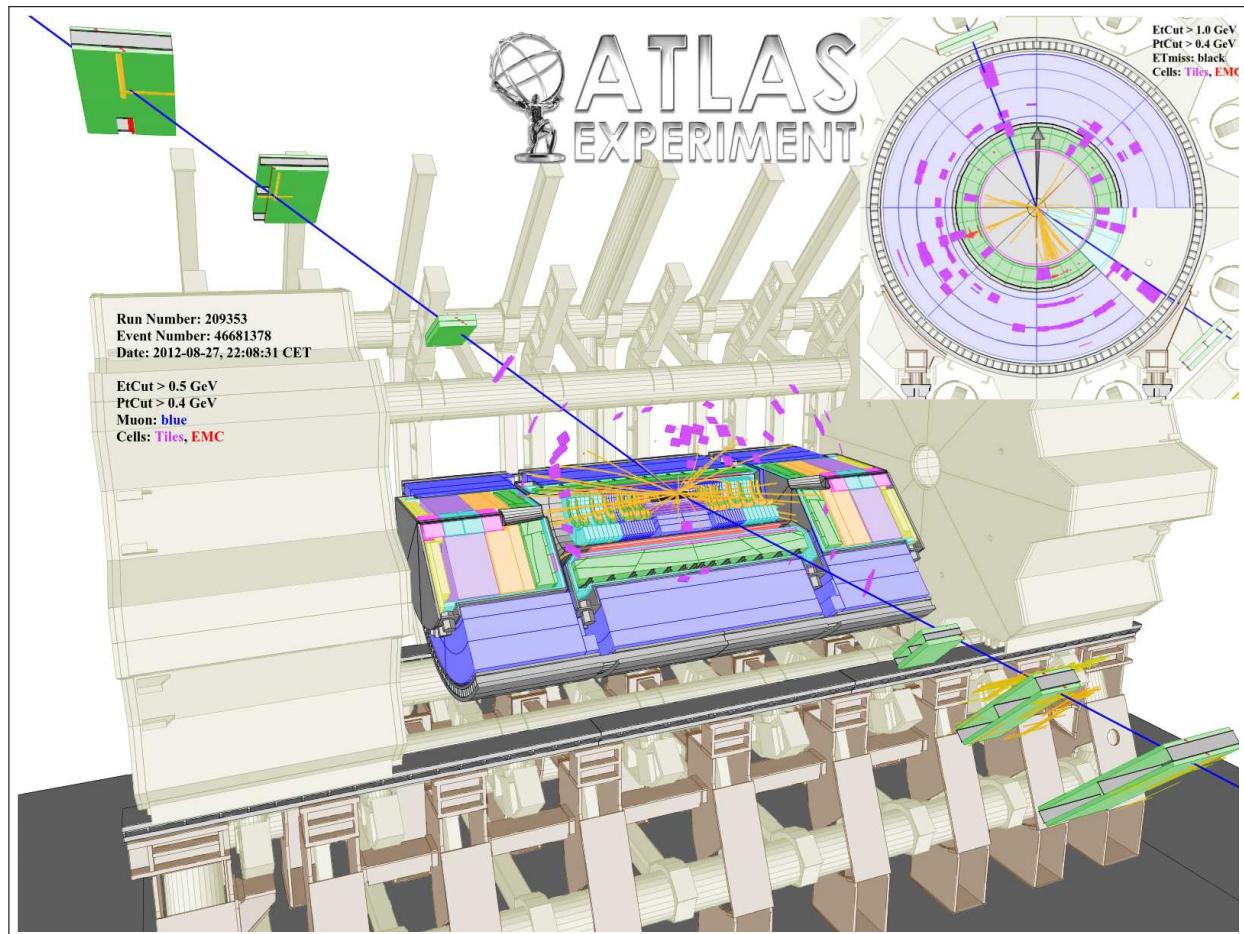


# Studying the Z Boson with the ATLAS Detector at the LHC

Version 1.2 (December 11, 2020)



Original Authors: Falk Bartels, Julia I. Djupsland, Manuel P. Geisler, Martin Wessels  
Contact: [martin.wessels@cern.ch](mailto:martin.wessels@cern.ch)

Advanced Physics Lab for Physicists  
Heidelberg University



# Contents

<b>1 Preliminary Remarks</b>	<b>3</b>
1.1 Brief Overview . . . . .	3
1.2 Prerequisites . . . . .	4
1.3 Preparation . . . . .	4
<b>2 Theoretical Foundations</b>	<b>5</b>
2.1 The Standard Model of Particle Physics . . . . .	5
2.2 Discovery of the W and Z Bosons . . . . .	6
2.3 The Drell-Yan Process at the LHC . . . . .	10
<b>3 The Large Hadron Collider and the ATLAS Experiment</b>	<b>12</b>
3.1 The Large Hadron Collider . . . . .	12
3.2 The ATLAS Experiment . . . . .	13
<b>4 Particle Identification</b>	<b>17</b>
4.1 Tracks and Clusters . . . . .	17
4.2 Electrons and Photons . . . . .	18
4.3 Muons . . . . .	19
<b>5 Practical Information and Tools</b>	<b>20</b>
5.1 Computing and Python Comments . . . . .	20
5.2 The Data Analysis Framework ROOT . . . . .	20
5.3 Monte Carlo Simulation . . . . .	21
5.4 Hadron Collider Event Kinematics . . . . .	23
5.5 The Tag-and-Probe Method . . . . .	25
5.6 ATLAS Open Data . . . . .	25
5.7 Uncertainties of the Measurement . . . . .	26
<b>6 Instructions</b>	<b>27</b>
6.1 Information about the Files in this Lab Course . . . . .	27
6.2 Getting to Know the Basics . . . . .	27
6.3 Automating Things . . . . .	28
6.4 Selecting Events . . . . .	29
6.5 Comparing Data and MC Simulation . . . . .	31

6.6	Fitting the Z Mass . . . . .	32
6.7	Determining Efficiencies . . . . .	33
6.8	Systematic Errors . . . . .	33
<b>Appendix</b>		<b>34</b>
<b>References</b>		<b>35</b>

# 1 Preliminary Remarks

In this lab course we will study the  $Z$  boson using proton-proton collision data recorded with the ATLAS detector at the CERN Large Hadron Collider, LHC. A set of real data and several supporting simulated Standard Model processes are available. Central techniques for performing a typical analysis in modern particle physics will be acquired. This lab course was first introduced in the winter semester 2018/19.

## 1.1 Brief Overview

The goal of this lab course is to gain an impression of a data analysis in modern particle physics. Particle physics is concerned with the fundamental constituents of the Universe, the elementary particles, and the interactions between them, the forces [1]. A modern particle physics experiment, such as the ATLAS experiment [2], requires years of design and planning, research and development, and finally construction of a detector the size of a five-story building. And of course it needs a particle accelerator such as the Large Hadron Collider, LHC, at CERN (Conseil Européen pour la Recherche Nucléaire – European Council for Nuclear Research) located astride the Franco-Swiss border near Geneva.

Due to the complexity of the machines and detectors, research in particle physics is performed in large collaborations. ATLAS for instance comprises of about 3000 scientific authors from 181 institutions around the world, representing 38 countries from all the world's populated continents.

The ATLAS collaboration has released an official dataset [3], [4], open to the public for educational purpose only, which will be used in this lab course. The dataset consists of real proton-proton collision data which were recorded with the ATLAS detector at the LHC in 2012. While these exact data were also part of the discovery of the famous Higgs boson that same year<sup>1</sup> the released statistics corresponding to an integrated luminosity of  $1 \text{ fb}^{-1}$  is not sufficient to extract the signal of the Higgs decay. Instead, we are using another boson for our investigations in this lab course, the  $Z$  boson.

The  $Z$  boson is a neutral elementary particle which – along with its electrically charged cousin, the  $W^\pm$  boson – carries the weak force. As with the Higgs boson, it was discovered at CERN, in 1983 at the Super Proton Synchrotron (SPS), operated as proton-antiproton collider. Owing to its electrical neutrality, the  $Z$  boson must decay into a particle-antiparticle pair. In this lab course we are making use of the clean detector signatures generated by  $Z$  bosons decaying into electron-positron and muon-antimuon pairs, as illustrated in figure 1. We will use the same methods and tools as used in a “real analysis” in order to extract these events out of the bulk of recorded data, reconstruct the  $Z$  boson candidates and determine its mass with fitting techniques.

Given this lab course is a data analysis, all work will be done with a computer. The main tasks will consist in the development and programming of analysis code to extract the  $Z$  candidate events, and to perform the various measurements and plots. We will be using the Python programming language, interfacing the data analysis package ROOT. Both belong to the standard tools in modern physics data analysis, not only in the field of particle physics.

<sup>1</sup> On 4 July, the ATLAS and CMS collaborations announced they had each observed a new particle in the mass region around 125 GeV consistent with a Higgs boson, the particle linked to the mechanism proposed in the 1960s to give mass to the  $W^\pm$ ,  $Z$  and other particles.

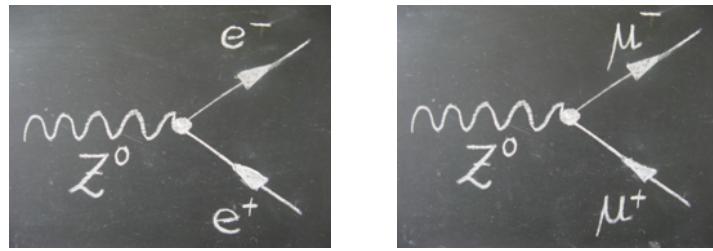


Figure 1: Leading order Feynman diagrams of the  $Z$  boson decay into electron and muon pairs [5].

## 1.2 Prerequisites

Basic knowledge in the field of particle physics is mandatory to participate in this lab course. The successful completion of the lecture and exercises on the experimental physics IV course on nuclear and particle physics (Experimentalphysik IV, PEP4) is required.

The main work in this lab course consists in the development of analysis code using the programming language Python. Therefore you should in general be interested in computer programming if you consider booking this course.

We require basic knowledge of the programming language Python to be present before starting the work. The course can in principle be booked without this prerequisite, but we expect you to acquire the missing knowledge elsewhere (also see subsection 1.3). Teaching how to program in Python is not covered in these instructions, but basic skills will be needed to successfully complete the course.

## 1.3 Preparation

In order to refresh your knowledge of the basic principles, the PEP4 lecture content on particle physics should be revisited, and in particular the properties of the  $Z$  boson should be familiarized. Since we will be studying proton-proton collision data recorded at the LHC, also scattering processes of quarks at highest energies play an important role. The overall functionality of the LHC and of the ATLAS detector needs to be reviewed.

We are using the programming language Python in this lab course. Basic programming skills in Python are assumed and cannot be taught within the limited time available in this lab course. It is required that you refresh or learn basic commands and coding elements in Python before starting this lab course.

An interactive introductory course into Python is offered within the framework of the Heidelberg University physics beginner's lab and can be found here:

<https://www.physi.uni-heidelberg.de/Einrichtungen/AP/Python.php> .

Another excellent tutorial is being provided by the Python Software Foundation itself and can be accessed through following link:

<https://docs.python.org/3/tutorial/index.html> .

In addition to the programming language Python we are using the data analysis package ROOT. “Getting Started” information and tutorials can be found online at <https://root.cern.ch>, including the (very) complete reference guide. While it will not hurt to skim through the pages of the **ROOT Primer** for preparation, most students learn ROOT “by doing” (see subsection 5.2).

## 2 Theoretical Foundations

A general introduction into the field of modern particle physics is given in the university's lecture on experimental physics IV (Experimentalphysik IV, PEP4) which is mandatory for this lab course. While the lecture provides all the physics background needed, this section repeats and discusses in more detail aspects which are of particular interest for this lab course.

After briefly recapitulating the basic principles of the Standard Model of particle physics, the subsequent subsections provide deeper insights into the subject of study itself, the  $Z$  boson. They are for the most parts based on two publications aiming at the general audience.

First, subsection 2.2 gives a historic review of the discovery of the  $W$  and  $Z$  bosons, based on an article written by Prof. P. M. Watkins, University of Birmingham, in 1986 [6]. The subsequent subsection 2.3 then discusses the Drell-Yan process at the LHC, which refers to the annihilation of a quark with an antiquark into a virtual photon or  $Z$  boson. Here the writing is inspired by an entry on the Quantum Diaries website [7] which from 2005 to 2016 hosted blogs by scientists from particle physics institutions around the world.

Needless to say that only short summaries of the most relevant and interesting parts of those articles can be given in this manual. Interested students however are invited to read the full articles which are publicly available, and details can be found in the reference section of this manual.

### 2.1 The Standard Model of Particle Physics

Today, the knowledge gained by scientists about the structure of matter is summarised in the Standard Model (SM) of particle physics, which has been developed in the second half of the last century and has proved to be very successful in describing a wide range of precise experimental results in the field of high energy physics<sup>2</sup>. The basic assumptions of the Standard Model are simple and can be written down in a few lines [9].

Matter is composed of elementary fermions with spin  $1/2$ , quarks and leptons, each occurring in three families consisting of two particles. Each of the 12 particles has a corresponding antiparticle with opposite charge but otherwise identical properties. A summary of the twelve fundamental particles of the Standard Model is given in table 1.

Leptons			Quarks		
Particle	Q	Mass [GeV]	Particle	Q	M [GeV]
electron ( $e^-$ )	-1	0.0005	down ( $d$ )	$-1/3$	0.003
neutrino ( $\nu_e$ )	0	$< 10^{-9}$	up ( $u$ )	$+2/3$	0.005
muon ( $\mu^-$ )	-1	0.106	strange ( $s$ )	$-1/3$	0.1
neutrino ( $\nu_\mu$ )	0	$< 10^{-9}$	charm ( $c$ )	$+2/3$	1.3
tau ( $\tau^-$ )	-1	1.78	bottom ( $b$ )	$-1/3$	4.5
neutrino ( $\nu_\tau$ )	0	$< 10^{-9}$	top ( $t$ )	$+2/3$	174

Table 1: The twelve fundamental particles of the SM and their masses (M) and charges (Q) in units of the elementary charge.

In the Standard Model, these elementary particles are subject to three fundamental forces, which are

<sup>2</sup> Despite its remarkable success in describing experimental results, the SM is just a model and there are many unanswered questions. An interesting outline of the physics reasons to pursue a future programme in high-energy colliders is amongst others presented by the Italian theoretical physicist G.F Giudice in [8].

the strong, the electromagnetic and the weak force. Gravity, the forth known fundamental interaction, is not included in the Standard Model. But given its small strength, the effects of gravitation can usually be neglected in particle physics experiments.

The interactions between the particles composing matter are mediated by bosons carrying spin 1. These exchange-particles, or force carriers, are the massless photon for the electromagnetic force, the massive  $Z$  and  $W^\pm$  bosons for the weak force, and eight massless gluons mediating the strong interaction. A summary is shown in table 2. The third kind of particle in the Standard Model is the Higgs boson, which is responsible for the creation of particle masses.

Boson	Mediated Force	Q	Spin	M [GeV]
Gluon ( $g$ )	Strong	0	1	0
Photon ( $\gamma$ )	Electromagnetism	0	1	0
$W$ Bosons ( $W^\pm$ )	Weak	$\pm 1$	1	80.4
$Z$ Boson ( $Z$ )		0	1	91.2

Table 2: The four force-carrying bosons of the SM.

Towards a grand unification of the fundamental forces, the electromagnetic and the weak force have been combined into the electroweak interaction (see subsection 2.2). The field theory describing the strong interaction is called Quantum Chromodynamics.

## 2.2 Discovery of the $W$ and $Z$ Bosons

The prediction as well as the discovery of the  $W$  and  $Z$  bosons were two remarkable achievements in theoretical and experimental physics, respectively, each honoured with its own Nobel Prize.

In 1979, the Nobel Prize in Physics was awarded jointly to Sheldon Lee Glashow, Abdus Salam and Steven Weinberg “for their contributions to the theory of the unified weak and electromagnetic interaction between elementary particles, including, inter alia, the prediction of the weak neutral current” [10]. Five years later, in 1984, the Nobel Prize in Physics was awarded jointly to Carlo Rubbia and Simon van der Meer “for their decisive contributions to the large project, which led to the discovery of the field particles  $W$  and  $Z$ , communicators of weak interaction” [11]. This section gives a historic review of the discovery of the  $W$  and  $Z$  bosons [6].

We currently believe that there are just those four fundamental forces or interactions – the gravitational and electromagnetic force as well as the weak and strong nuclear forces – which all however have very different properties. For example, the electromagnetic force acts only between particles carrying an electric charge, and obeys the inverse square law with long range. Contrary, the weak force, which is responsible for the  $\beta$ -decay and for many of the nuclear interactions fuelling the energy generation of the sun and stars, acts between all types of particles, but only over very small distances, of approximately  $10^{-17}$  m.

On first sight, the weak nuclear force seems to have little in common with electromagnetism. However it was precisely the previously mentioned unification of both forces into a combined theory which introduced the massive  $W^\pm$  and  $Z$  bosons as mediators of the weak force, and successfully predicted their properties.

While the abstract concept of a field is used in classical physics in order to describe the (electromagnetic) interaction between particles, in particle physics a more convenient model is introduced, in which the exchange of a particle transmits the force between the interacting particles<sup>3</sup>. These so-

<sup>3</sup> Given that Quantum Field Theory (QFT) underlies modern elementary particle physics, also the field concept plays of

called virtual particles are created on borrowed energy, which is allowed for a short time, by Heisenberg's Uncertainty Principle. The uncertainty principle also allows to estimate the typical range of a force carried by a virtual particle to depend inversely on the mass of the force carrier. The virtual photon has a mass near zero, and hence the range of the electromagnetic force is infinite. Clearly much heavier carriers are needed for the short-ranged weak interaction; these are the  $W^\pm$  and  $Z$  bosons with masses about one hundred times heavier than that of a proton.

The properties of the weak interaction have been studied in the middle of the last century for many years at experiments at high energy accelerators around the world. This can be done most effectively in reactions where the other stronger interactions cannot participate. Since leptons do not feel the strong force, for example the decay of a muon to an electron must proceed by the weak interaction. And since neutrinos are electrically neutral (leptons), in general all interactions involving neutrinos must proceed via the weak force (or gravity).

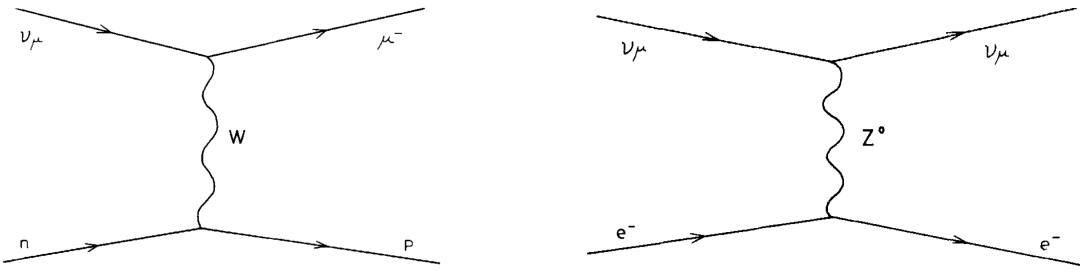


Figure 2: Leading order diagrams for the charged current interaction between a neutron and a neutrino (left), and for a neutral current interaction between an electron and a neutrino (right).

The left diagram in figure 2 shows the weak charged current interaction between a neutron and a neutrino where a  $W$  boson is exchanged. Every particle carries a weak charge (not to be confused with the electric charge), and the  $W$  acts between these charges as the carrier of the weak force. In the interaction, the neutron and neutrino in the initial state convert into a proton and muon, respectively<sup>4</sup>. This change of the particle species, usually referred to as particle flavour, is not only characteristic but unique for the weak charged current interaction.

At this time all detailed studies of neutrino interactions suggested that the neutrino always changed its identity in any collision, which means that a charged  $W$  boson must have mediated the force. However, collisions detected in a bubble chamber exposed to a neutrino beam at CERN in 1973, could only be explained by the exchange of a neutral force carrier, such as the  $Z$  boson. In these so-called neutral current interactions, the neutrino collided with the target yet remained as neutrino, as depicted in the right diagram in figure 2. This observation was a crucial result in attempts to unify the fundamental forces.

But how could the long range electromagnetic force be unified with the short range weak nuclear force? After the discovery of the weak neutral currents, there was now at least one common feature, which is the transmission of a neutral force carrier. However, detailed predictions were only possible when the gauge principle was applied.

All successful Quantum Field Theories of fundamental forces satisfy *local gauge invariance* which means that no observable effect is produced by any changes to the phase of particles, even if different changes are made at all points in space and time. This very restrictive requirement means that the properties of the force carriers are highly constraint, and the electroweak theory, developed by S.

course a very important role.

<sup>4</sup> For the conversion of the neutron to proton, the underlying process is the conversion of one of the initial state down-quarks into an up-quark when interacting with the  $W$  boson.

Glashow, A. Salam and S. Weinberg with the attempt to unify the electromagnetic and weak forces, was able to predict the properties of the  $W^+$ ,  $W^-$  and  $Z$  boson very precisely.

Contrary to the electromagnetic force, two charges are involved in the weak nuclear force, known as the weak isospin and the weak hypercharge<sup>5</sup>. The three carriers of the weak force related to the weak isospin are known as  $W^+$ ,  $W^-$  and  $W^0$  while the force carrier related to the weak hypercharge is called the  $B^0$  boson. The deep connection between the electromagnetic and the weak force is provided by the  $W^0$  and  $B^0$  bosons, which do not appear in nature but they mix together to produce the observable photon and  $Z$  boson.

The strength of the mixing is controlled by the parameter  $\theta_W$  known as the Weinberg angle. Its magnitude is not predicted by the electroweak theory, however it can be measured in many different experiments, and once known, the masses of the physical  $W^\pm$  and  $Z$  can be predicted. The first numerical values so obtained ranged from 60 to 80 GeV for the  $W$  mass, and from 75 to 92 GeV for the  $Z$  mass, too large to be accessible by any accelerator in operation at that time.

In 1976 a new idea, which would enable much higher energy collisions to be studied, was suggested by D. Cline, P. McIntyre and C. Rubbia. This proposed the transformation of an existing high-energy proton accelerator into a proton-antiproton collider as a quick and relatively cheap way to achieve collisions above threshold for  $W$  and  $Z$  production. Antiprotons have the opposite charge to protons, and so if injected into the same synchrotron as the protons but in opposite direction, the same electric fields can be used to accelerate these counter rotating beams simultaneously.

After careful evaluation of the most difficult aspects, related to creation of the antiprotons, beam compression and civil engineering of tunnel and experiments, the CERN laboratory approved the proton-antiproton collider project in 1978<sup>6</sup>.

To achieve high energetic collisions of antiprotons and protons, the existing super proton synchrotron (SPS) at CERN was used. In normal operation, the proton synchrotron (PS) accelerated protons to 26 GeV and these were injected into the SPS, where they were further accelerated to 400 GeV and ejected onto external fixed targets. In collider operation however, antiprotons and protons were first separately filled into the PS and pre-accelerated to 26 GeV, and then injected from the PS into the SPS, where both beams were accelerated to 270 GeV, resulting in an available collision (centre-of-mass) energy of 540 GeV. These beams were usually maintained in stable orbits for over 24h. At two points, where the beams intersected, large detectors were installed, known as UA1 and UA2 (Underground Area). In these regions the beams were focussed to even smaller cross sections, to increase the chance of antiproton-proton collisions.

The SPS started to operate as antiproton-proton collider in 1981 (until 1991), as such it was also called Spp̄S. In the early operation, many difficulties had to be overcome in this complex project, which however provided useful experience in commissioning the new detectors. Early in October 1982, the first serious candidates for  $W$  boson production were identified and by the end of that year, both the UA1 and UA2 experiments had isolated a handful of  $W$  candidate events. Intense analysis showed that they were consistent, both in production rate and decay characteristics, with the first ever observation of  $W^+$  and  $W^-$  bosons. On 25 January 1983 a press conference was called at CERN to formally announce the discovery of the  $W^\pm$  intermediate vector bosons, the carriers of the weak nuclear force.

<sup>5</sup> In the GSW model, the underlying gauge symmetry is the  $U(1)_Y$  of weak hypercharge, and  $SU(2)_L$  of weak isospin, written as  $U(1)_Y \times SU(2)_L$ .

<sup>6</sup> For example, the production of a useful antiproton beam took one day to achieve, which was collected until further use in the so-called antiproton accumulator (AA), a ring of magnets with a diameter of 50 m. Also the required beam compression, or cooling, could only be realised using “stochastic cooling”, a new method which was invented by S. Van der Meer, an accelerator physicist at CERN, in 1968.

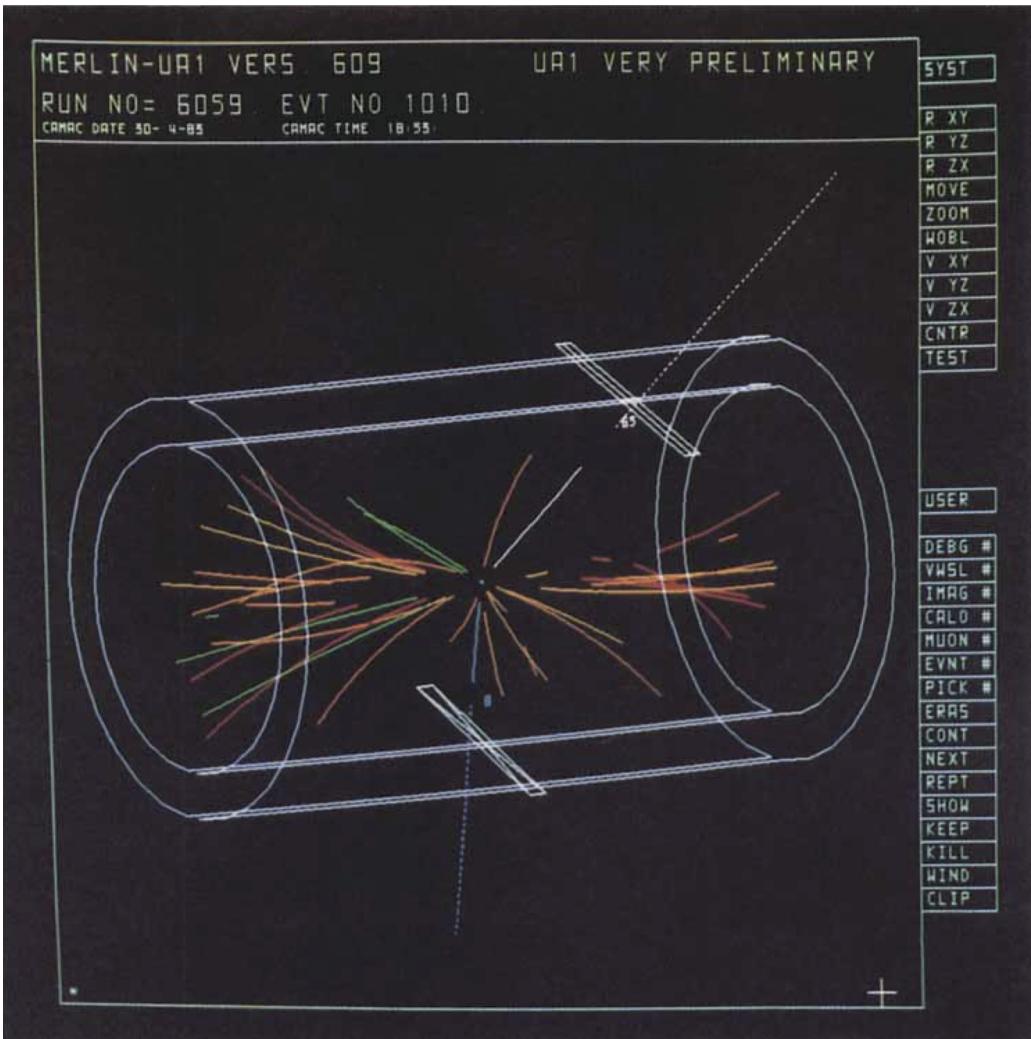


Figure 3: A general view of the first  $Z$  boson candidate event detected in the UA1 experiment. To reduce complexity of the display, tracks and calorimeter cells are shown only if their transverse momenta exceed 1 GeV. The two most energetic tracks (blue and white) deposit very large energies at their entry to the electromagnetic calorimeter (large rectangles) and are consistent with the decay of the  $Z$  boson into an electron-positron pair.

The search for the  $Z$  boson was the main objective of the next collider run, starting in April 1983. Despite the decay modes of the  $Z$  boson into electron-positron or muon-antimuon pairs being much simpler to identify than those for the  $W$  decay with its elusive neutrino, the production rate of  $Z$  bosons was even smaller than that of  $W$  bosons. Nonetheless, after having raised the beam intensities, the first candidate event for a  $Z$  decay to an electron-positron pair was found by the UA1 experiment on 4 May 1983. It had been recorded by the detector on the evening of the 30 April, and is shown in figure 3. A few days later, also the first candidate event for the decay to a muon-antimuon pair was identified by the UA1 experiment.

On 1 June 1983, CERN formally announced the discovery of the  $Z$  boson. Towards the end of the data taking period in July 1983, the UA1 and UA2 experiments had collected nine  $Z$  bosons and were able to quote a preliminary mass which was in excellent agreement with the prediction from the electroweak theory.

While the  $Z$  and  $W^\pm$  bosons were successfully discovered in proton-antiproton collisions, the ideal

machine to produce the weak bosons and to measure their properties in the most convenient experimental conditions is an  $e^+e^-$  collider, as beautifully demonstrated by the success of the Large Electron Positron (LEP) collider program at CERN. The machine operated in two phases. In the first phase, between 1989 and 1995, 18 million  $Z$  bosons were collected, while in the second phase, from 1996 to 2000, some 80 thousand  $W$  bosons were generated at energies gradually climbing from the  $W$ -pair threshold to the maximum of 209 GeV [12].

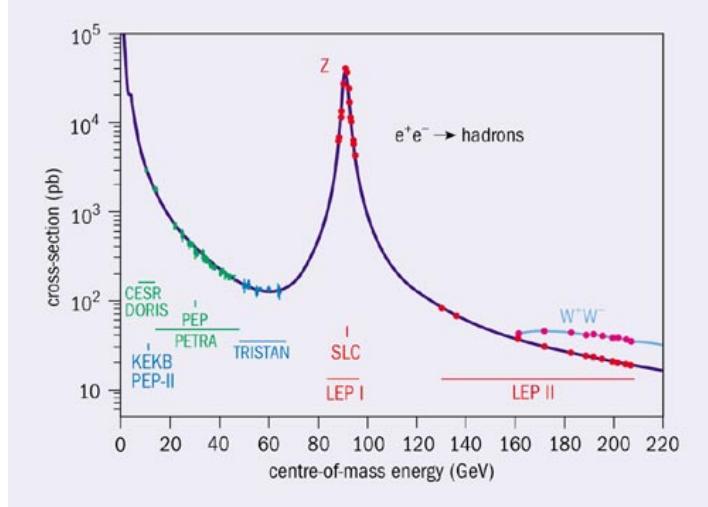


Figure 4: The  $e^+e^-$  annihilation cross-section to hadrons, from initial low energies in early colliders to the maximum energy at LEP.

Outstandingly clear events could be observed in each of the four LEP experiments, OPAL, DELPHI, L3 and ALEPH. As a result, the experimental analysis of the formation cross-section as depicted in figure 4, i.e. the line-shape parametrized in Breit-Wigner form by the  $Z$  mass and its width, could be performed with precision unprecedented in high-energy experiments, obtaining values for the mass  $M_Z = 91187.5 \pm 2.1$  MeV and width  $\Gamma_Z = 2495.2 \pm 2.3$  MeV.

### 2.3 The Drell-Yan Process at the LHC

Particle colliders such as the Large Hadron Collider are, in a sense, very powerful microscopes [7]. The higher the collision energy, the smaller distances can be studied. One of the most important processes that occurs in proton-proton collisions is the Drell-Yan process.

When a quark (e.g. a down quark  $d$ ) from one proton collides with an antiquark of the same flavour (e.g. an antidown quark  $\bar{d}$ ) from the oncoming proton, they can annihilate into a virtual photon  $\gamma^*$  or a  $Z$  boson if the net electric charge is zero<sup>7</sup>. After briefly propagating, the  $\gamma^*/Z$  can split into a lepton and its antiparticle partner, for example into an electron-positron pair or into a muon and antimuon. In pictures, the annihilation of a  $d$  quark and  $\bar{d}$  antiquark into a muon-antimuon pair looks like figure 5a.

By conservation of momentum the sum of the muon and antimuon momenta will add up to the momentum of the  $\gamma^*/Z$ . This is nicely illustrated in figure 6 where the invariant mass distribution for any opposite-sign muon pair produced in proton collisions at the 7 TeV LHC is plotted, as measured by the CMS experiment.

The rightmost peak at about 90 GeV corresponds to the production of  $Z$  bosons. The other peaks

<sup>7</sup> If the net electric charge of the two incoming quarks is 1 (e.g. for  $u, \bar{d}$ ) or  $-1$  (e.g. for  $\bar{u}, d$ ) a  $W^+$  boson or  $W^-$  boson is produced instead of the electrically neutral photon or  $Z$  boson.

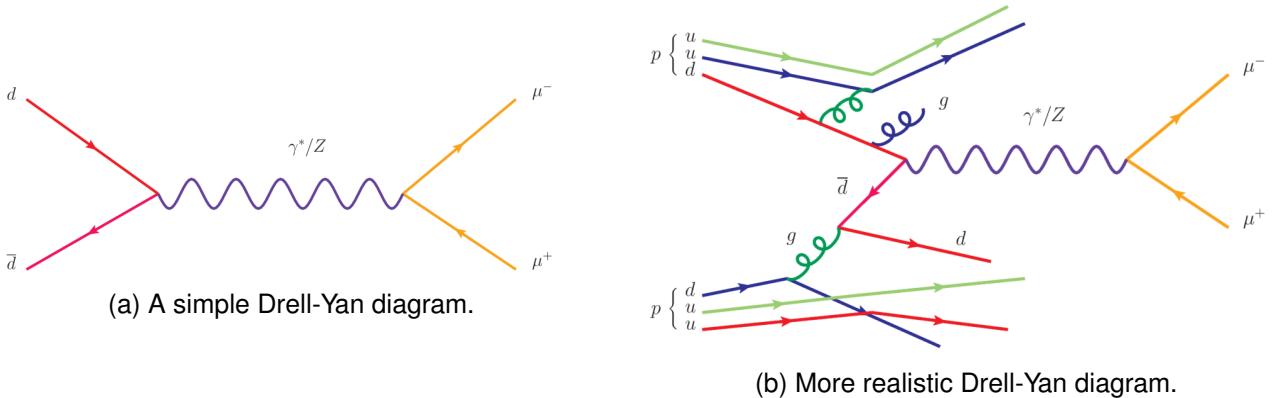


Figure 5: Diagrams for the Drell-Yan process at the LHC.

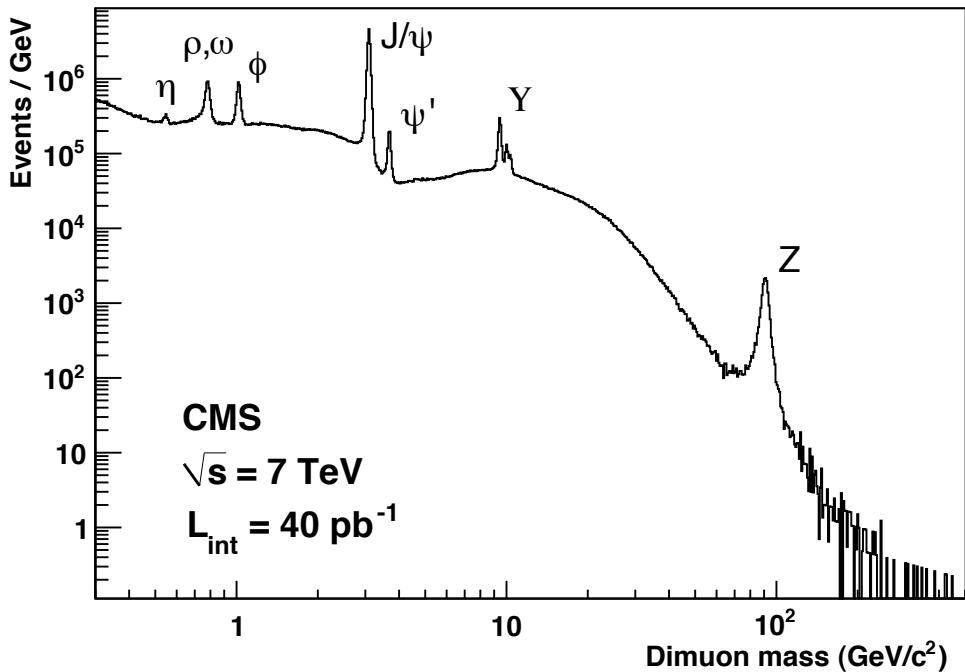


Figure 6: CMS invariant-mass spectra of opposite-sign muon pairs using 2010 data [13].

represent the production of similarly well-known particles in the particle zoo that have decayed into a muon-antimuon pair. The clarity of each peak and the fact that this plot uses only about 0.2% of the total data collected during the first LHC data collection period (Run 1) means that the Drell-Yan process is a very useful tool for calibrating the detectors. If the experiments are able to measure the  $Z$  boson, the  $\rho$  meson, etc. at their correct energies, then there is confidence that the detectors are working well enough to study nature at energies never before explored in a laboratory.

However, the Drell-Yan process is not as simple as drawn in figure 5a. Real collisions include the remnants of the scattered protons. An example for a more realistic Drell-Yan interaction is depicted in figure 5b, in which in addition to the proton remnants gluon radiation processes are included. This will lead to more particles in the final state, in particular jets of hadrons produced by the hadronization of quarks and gluons.

### 3 The Large Hadron Collider and the ATLAS Experiment

This section gives an overview over the Large Hadron Collider and the ATLAS experiment. Those components of the ATLAS detector which are of particular interest for this lab course are explained in greater detail.

#### 3.1 The Large Hadron Collider

The Large Hadron Collider is the world's current most powerful particle collider. It is located up to 175 m underground at the European Centre for Nuclear Research (CERN) in Geneva, Switzerland. Inside the accelerator ring with its almost 27 km circumference and state-of-the-art magnets, protons<sup>8</sup> are accelerated to energies<sup>9</sup> of  $\sqrt{s} = 13 \text{ TeV}$ . The design luminosity of the LHC is  $10^{34} \text{ cm}^{-2} \text{ s}^{-1}$ .

A chain of accelerators, mostly stemming from previous CERN experiments, pre-accelerates the protons stepwise up to 450 GeV before injecting them into the LHC. The LHC hosts two beams in separate beam pipes, one circulating clockwise the other counter-clockwise. At interaction points (IPs) the beams are crossed, particles collide and detectors measure the interaction processes.

The colliding beams are structured into bunches. Each beam ( $i = 1, 2$ ) has 3564 possible bunch slots, of which in maximum  $n_b = 2808$  can be filled with protons. Every filled bunch slot contains typically  $N_i = \mathcal{O}(10^{11})$  protons. Successive bunch slots are spaced in time by 25 ns, which is a result of the circumference of the ring, the total number of bunch slots and the speed of light. The revolution frequency  $f_{\text{rev}}$  of a bunch around the whole ring is 11.25 kHz. A single bunch is focussed to  $\approx 30 \text{ cm}$  in length which corresponds to  $\approx 1 \text{ ns}$  in time.

The so-called lattice of the LHC, which refers to the setup of the optical elements (electromagnets), uses high-end technologies. Superconducting magnets with a temperature of roughly 2 K are used to keep the protons on track. The large energies involved require dipole magnets of up to 8 T and electrical currents of about 11 kA. Beam bending is accomplished by 1232 dipole magnets, beam focusing by 382 quadrupole magnets, and 3700 additional magnets take care of various further tasks such as the beam dump at the end of a fill.

Four main experiments are located around different IPs: ATLAS and CMS are multi-purpose experiments with a broad physics program ranging from precision measurements of Standard Model interactions to searches for Dark Matter. LHCb is a smaller experiment specialized in measurements of the flavour sector of the SM. Lastly, ALICE focuses on the mode of operation when heavy ions rather than protons are colliding in the LHC. A schematic illustration of the accelerator ring and the four experiments can be found in figure 7.

A quantity of special relevance for collider experiments is the instantaneous luminosity  $\mathcal{L}$ , which is a measure for the ability of a particle accelerator to produce the desired number of collisions. It is defined as

$$\mathcal{L} = \frac{N_1 N_2 f_{\text{rev}} n_b}{4\pi \sigma_x \sigma_y} . \quad (1)$$

The subscripts 1 and 2 refer to the two beams.  $\sigma_x$  and  $\sigma_y$  denote the horizontal (x) and vertical (y)

<sup>8</sup> Heavy ions can be injected into the machine, too. For this lab course these are not used.

<sup>9</sup> This energy was achieved during Run-2 of the LHC, which lasted from 2015 to 2018. In the first operational LHC Run-1 between 2009 and 2012, energies of up to  $\sqrt{s} = 8 \text{ TeV}$  were reached. The design centre-of-mass energy is 14 TeV which is expected to be achieved in the future Run-3 (foreseen start in mid 2021).

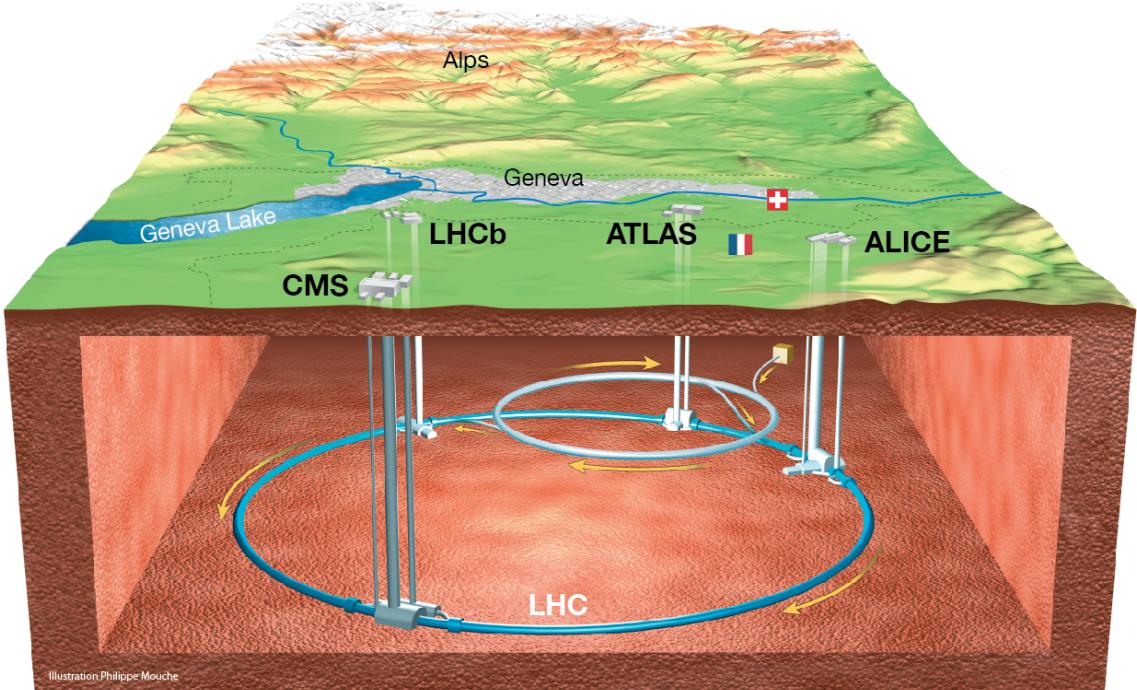


Figure 7: A schematic view of the LHC ring with the four main experiments in the Geneva region.

beam spread and are a measure for how focused the beams are<sup>10</sup>. A related quantity is the integrated luminosity, which yields the size of a collected data set. It is defined as

$$\mathcal{L}_{int} = \int \mathcal{L} dt . \quad (2)$$

The number of events  $N$  which involve a specific transition  $pp \rightarrow X$  of protons into a final state  $X$ , for instance a  $Z$  boson or a Higgs boson, in a given data set with  $\mathcal{L}_{int}$  can be calculated via

$$N = \sigma_{pp \rightarrow X} \cdot \mathcal{L}_{int} , \quad (3)$$

where  $\sigma_{pp \rightarrow X}$  is the production cross-section for the final state  $X$  in proton-proton collisions.

### 3.2 The ATLAS Experiment

ATLAS is an approximately cylindrically shaped, huge general-purpose detector with a length of 44 m, a width of 25 m and a mass of 7000 t. It consists of three main subdetectors in an onion-like arrangement around the beam pipe, and two magnetic systems. The latter bend trajectories of electrically charged particles and allow for the determination of their momenta and charges.

The three subdetectors are the so-called inner detector for tracking which is located closest to the beam pipe, a calorimeter system surrounding the tracker for energy measurements and shower

<sup>10</sup> This formulation assumes Gaussian intensity profiles of the beams.

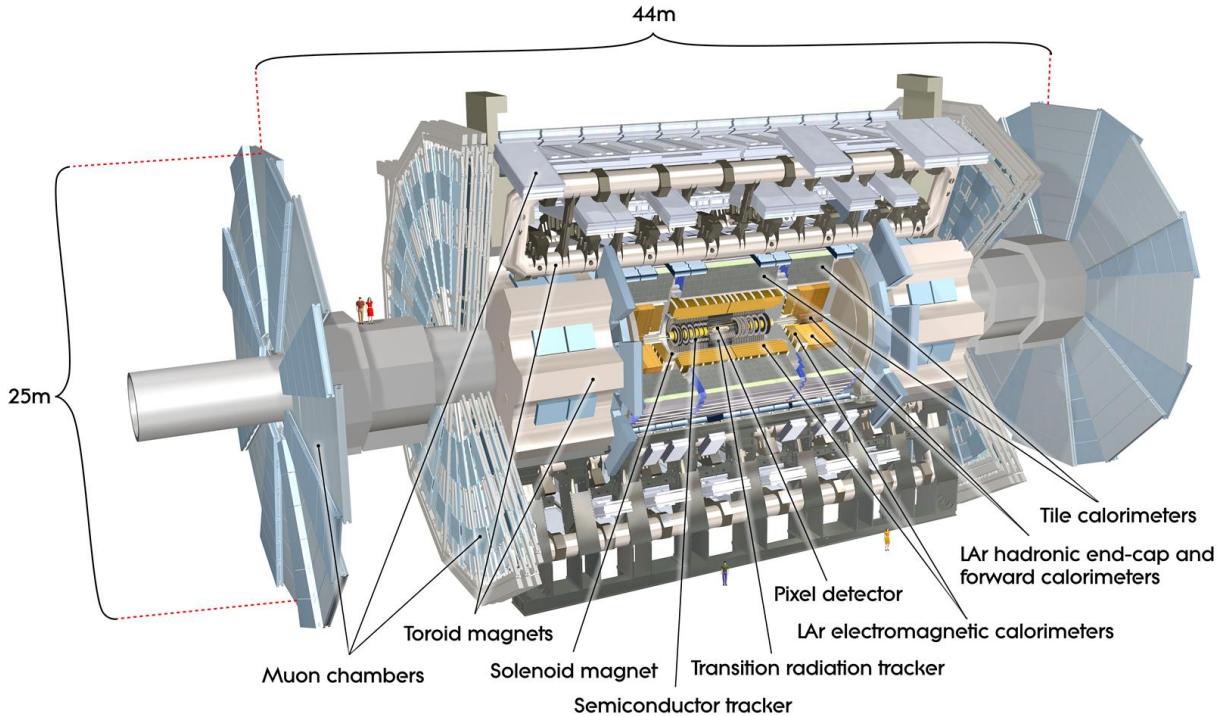


Figure 8: A schematic view of the ATLAS detector.

shape recognition, and a muon spectrometer in farthest distance to the beam pipe. A schematic view is depicted in figure 8. The subdetectors are described in more detail in the following.

### The Inner Detector

The inner detector's (ID) main objective is the reconstruction of trajectories (or tracks) of electrically charged particles. It yields important information for particle identification and the measurement of their transverse momenta  $p_T$ . It also plays a crucial role in the reconstruction of primary and secondary vertices and thus helps in assigning particles to their origin. It measures the energy loss of charged particles traversing a medium via ionization as described by the Bethe-Bloch equation. Various types of tracking systems identify the products of this ionization, i.e. either electrons or electron-hole pairs, or make use of transition radiation.

The ID as a whole can reliably reconstruct tracks of charged particles with a  $p_T > 400$  MeV. It covers a range of  $|\eta| < 2.5^{11}$  and consists of three tracking detectors, themselves arranged in an onion-like structure. From innermost to outermost these are the pixel detector (Pixels), the semi-conductor tracker (SCT), and the transition radiation tracker (TRT). The radial coverages of each are 3-15 cm (Pixels), 30-56 cm (SCT), and 56-107 cm (TRT).

<sup>11</sup> ATLAS uses a right-handed coordinate system with its origin at the nominal interaction point (IP) in the center of the detector and the  $z$ -axis along the beam pipe. The  $x$ -axis points from the IP to the center of the LHC ring, and the  $y$ -axis points upwards. Cylindrical coordinates  $(r, \phi)$  are used in the transverse plane,  $\phi$  being the azimuthal angle around the  $z$ -axis. The pseudorapidity is defined in terms of the polar angle  $\theta$  as  $\eta = -\ln \tan(\theta/2)$ . Angular distance is measured in units of  $\Delta R \equiv \sqrt{(\Delta\eta)^2 + (\Delta\phi)^2}$ .

## The Calorimeters

The ATLAS calorimeter consists of two main components, the electromagnetic (ECAL) and the hadronic calorimeter (HCAL). They are designed to fully contain the showers of electromagnetically and strongly interacting particles, providing precise energy measurements. In addition, the fine segmentation of the calorimeters into cells allows for the reconstruction of shower shapes, which aids the discrimination between different kinds of particles and hence improves particle identification. ATLAS uses sampling calorimeters, meaning that the calorimeters are made of two kinds of materials sandwiching each other: an active material in which the energy measurement takes place, and a dense absorbing material which fuels particle showering and leads to containment of the shower. The whole calorimeter system covers the range  $|\eta| < 4.9$ .

The ECAL uses liquid argon (LAr) as active material and lead for the absorbers. Ionization takes place in the active material and copper-etched electrodes extract the resulting electrical signals. For the HCAL, LAr technology is used in the forward and backward end-cap, while the central barrel component is utilizing scintillating tiles as active material and steel for the absorber (Tile). The active material measures the nuclear interactions that take place during hadronic showers.

## The Muon Spectrometer

High energetic muons are, apart from neutrinos, the only particles that can penetrate through the HCAL and escape the ATLAS detector. Being electrically charged they leave products of ionization. Hence the muon spectrometer (MS) is an additional tracking system enveloping the rest of the detector to allow for the extended reconstructions of muon tracks. Several types of so-called muon chambers exist which measure spatial and temporal components of muon tracks using different techniques and various ionization gases. Muon tracks can be reconstructed within the acceptance  $|\eta| < 2.7$ .

## The Magnet System

ATLAS features a unique hybrid system of four large superconducting magnets. This magnetic system is 22 m in diameter and 26 m in length, providing the magnetic field over a volume of approximately 12,000 m<sup>3</sup>, with a stored energy of 1.6 GJ. It consists of one solenoid and three toroids (one barrel and two end-caps). The solenoid is aligned on the beam axis and provides a 2 T axial magnetic field for the inner detector, while minimising the radiative thickness in front of the barrel ECAL. The barrel and two end-cap toroids produce a toroidal magnetic field of approximately 0.5 T and 1 T for the muon detectors in the central and end-cap regions, respectively.

## Trigger and Data Acquisition

With a collision frequency of 40 MHz and a raw event size in the order of 1 MB, the bandwidth of the ATLAS detector as well as the storage capacities are neither sufficient to read out the full event information for every bunch-crossing nor to store the results. Hence, the ATLAS detector is equipped with a trigger system which has the essential task to decide on the fly which bunch-crossing contains an interesting event and is to be read out and saved for future analyses. The trigger system consists of two stages: the custom hardware based Level-1 Trigger and the High Level Trigger (HLT) running on commercial off-the-shelf computer components with demanding requirements on network data flow and management.

The Level-1 trigger has the task to reduce the bunch-crossing rate of up to 40 MHz – corresponding to the 25 ns bunch spacing – to a rate of approximately 100 kHz. This decision has to be taken in less than  $2.5 \mu\text{s}$  because the readout buffers on the detector front-end electronics might otherwise overflow<sup>12</sup>. Hence signals from the calorimeters and the muon system are processed at reduced granularity in the corresponding Level-1 trigger subsystems, L1Calo and L1Muon. Handling the full calorimeter cell resolution or including information from the tracking detectors cannot be achieved within the  $2.5 \mu\text{s}$  time budget.

If particle candidates exceeding certain energy thresholds are identified, corresponding multiplicity bits are sent to the Central Trigger Processor (CTP) which, depending on a so-called trigger menu, might issue a L1-Accept (L1A) signal initiating the detector readout. In addition information about the location is transferred as a Region of Interest to the HLT.

Upon receipt of a L1A signal, the HLT executes reconstruction algorithms based on the full calorimeter granularity, including information from the high precision muon chambers and inner tracker. In order to save computation time, most of these algorithms operate only on the Regions of Interest identified by their corresponding Level-1 seed. This way the L1 acceptance rate of approximately 100 kHz is reduced to the final trigger rate of  $\approx 1 \text{ kHz}$  for events to be stored permanently for physics analysis.

---

<sup>12</sup> In fact the length of the buffers and the allowed latency of the Level-1 trigger were defined in interplay, taking into account performance needs balancing technical possibilities at the time ATLAS was designed.

## 4 Particle Identification

Particles are reconstructed and identified based on the signatures and energy deposits they leave in the various subdetectors of the experiment. Figure 9 shows the typical layout of a large particle physics detector such as ATLAS, and the corresponding signatures produced by different particles.

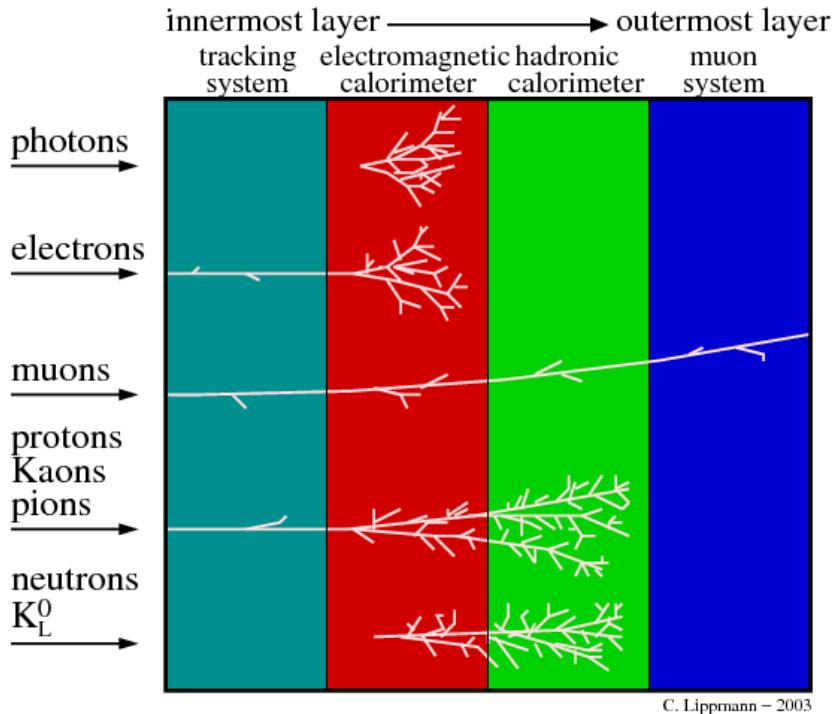


Figure 9: A schematic view of the traces various particles leave in different parts of a multi-purpose detector [14].

In general three main steps are involved in particle detection: reconstruction, calibration, and identification. Reconstruction algorithms yield a sample of candidates for particles types, for instance electron candidates, based on typical signatures of these particles. Calibration algorithms correct the measured quantities to their appropriate scales depending on the type of particle. However, the sample of particle candidates after reconstruction and calibration is typically very impure. For instance, the sample of calibrated electron candidates also contains many photons and hadronic jets. Hence the last step, identification, takes care of curating the sample by dismissing particle candidates which likely have a different origin. For this often multivariate techniques such as neural networks or boosted decision trees are used.

Broadly speaking, two kinds of information play the central roles for particle detection: tracking information from the ID and MS, and calorimeter clusters. Hence, before briefly discussing in subsections 4.2 and 4.3 the reconstruction of particles relevant to this lab course, a short overview on tracks and clusters is provided.

### 4.1 Tracks and Clusters

**Inner Detector tracks** are created by electrically charged particles which ionize the active material of sensor modules. Two kinds of algorithms exist for their reconstruction: an inside-out and an outside-in algorithm.

The inside-out algorithm aims at tracks from primary particles, defined as final state particles with a mean-lifetime  $t_p > 3 \cdot 10^{-11}$  s. Its starting point are three-hit seeds in the SCT and Pixel detector. After resolving potential track ambiguities it extends the track candidate outwards into the TRT. The outside-in algorithm's focus is finding tracks of secondary particles, which are produced in interactions of primary particles and typically have an origin different from the primary vertex of the event. As the name suggests, the algorithm starts with the outermost subdetector of the ID and moves inwards.

Reconstructed tracks have an associated *quality*. The quality of a track depends on how many total hits in all parts of the ID it has (the more the better) and on how well a  $\chi^2$  fit can be performed. The higher (tighter) the quality, the more certain the track resulted from ionization of a charged particle that originated from a relevant vertex.

Reconstruction of **Muon Spectrometer tracks** follows a similar general idea but a different algorithm is used. Here the algorithm starts with muon track candidates in the middle layer of the MS. Additional segments from other chambers inwards or outwards are added based on hit multiplicity, angles, positions and fit quality.

**Calorimeter clusters** are built from individual calorimeter cells and are indicative of energy losses of particles in the active medium. There are two main algorithms ATLAS uses for clustering cells: the sliding window algorithm and the topological algorithm. Electron and photon candidates are built from the sliding window algorithm. Hadronic jets are built from topological clustering. Since electrons play a central role for this lab course, the former algorithm will be briefly described in the following.

The sliding window algorithm divides the calorimeter into  $200 \times 256$  calorimeter towers in the  $\eta - \phi$  plane. A window of fixed size, for instance  $3 \times 5$  calorimeter towers, is slid across the entire plane. For each possible window position the energies of the constituent towers are summed. If a local energy maximum is found and if it is above a certain noise threshold, a so-called seed for a cluster is identified. This seed is turned into a final cluster by iterating over all calorimeter layers and by including additional cells within layer-dependent windows. The position of the cluster is defined as the energy-weighted barycentre of the cluster.

## 4.2 Electrons and Photons

Electron reconstruction starts with calorimeter clusters. The size for the sliding window algorithm as described in subsection 4.1 is  $3 \times 5$  towers, each with a size of  $0.025 \times 0.025$  in  $\eta - \phi$ . This cluster reconstruction yields a high reconstruction efficiency for electrons: 95% (99%) for electrons with  $p_T$  of 7 GeV (15 GeV).

Since electrons carry an electric charge, they also leave tracks in the ID. Tracks that might belong to the cluster are fit with an electron hypothesis, and if the extrapolated track points to the cluster, the two are matched. In the case of multiple valid tracks, the ambiguity is resolved using additional information such as the track-cluster distance or the number of hits of the tracks in the Pixel detector.

Usually the energy measurement of the electron is based on the cluster information whereas the position measurement is based on the track. This dependency of the electron reconstruction on tracking information means that it can only be performed in the region covered by the ID, i.e. within  $|\eta| < 2.47$ .

Electron candidates are calibrated to an appropriate electromagnetic energy scale. For this, a mixture of simulated data (cf. subsection 5.3) and measured reference events of the process  $pp \rightarrow Z \rightarrow e^+e^-$  are used.

Applying the basic reconstruction algorithm, the sample of electron candidates is still impure and an identification algorithm is required to resolve this. In order to discriminate real electrons from photons

or hadronic jets which accidentally left an electron-like detector signature, a likelihood (LH) method is used.

The LH method is a multivariate analysis technique that scores electron candidates according to certain metrics. Variables that are considered for this are for instance shower shape information using calorimeter clusters and track-cluster matching. Based on the LH score three working points are defined: *loose*, *medium*, and *tight*. An electron candidate labelled *tight* has the highest probability of corresponding to an electron. An electron candidate labelled *loose* on the other hand still has a fair chance of in reality corresponding to a photon or a hadronic jet.

In addition to the identification algorithm isolation requirements can be applied. An isolation requirement enforces that there is little to no other significant activity in close vicinity around the electron candidate. This requirement helps discriminating *prompt* electrons, such as in  $Z \rightarrow e^+e^-$  decays, against electrons originating for instance from photon conversions or the decays of heavy quarks.

Photon reconstruction, calibration, and identification follows a similar strategy as for electrons. However, due to their electrical neutrality photons do not leave tracks in the ID. Hence clusters without associated tracks are classified as photon candidates. In about 40% of the cases however, a photon conversion takes place and the photon converts into two electrons which hit the calorimeter in close vicinity. In this case two tracks can be found in the ID which often lack hits in the first layer.

### 4.3 Muons

Muon reconstruction relies on tracks both in the ID and the MS. Calorimeter information is usually not used as muons are minimum ionising particles and travel significant distances even in dense materials such as iron. Hence they do not leave large deposits in the calorimeters<sup>13</sup>.

In order to form muon candidates, tracks from the ID and MS are fit using a muon hypothesis. If the fit quality is good, a muon candidate is stored. If it is bad, a different combination of ID and MS tracks is fit. Calibration, identification, and isolation follow similar trains of thought as above.

---

<sup>13</sup> For example, a 10 GeV muon loses approximately  $13 \text{ MeV cm}^{-1}$  in iron and therefore has a range of several meters [1].

## 5 Practical Information and Tools

This section contains a collection of fairly unrelated topics. Practical information on computing programs and (software) tools will be given, as well as explanations of analysis methods, which will be made use of in this lab course. It is recommended to read it carefully in the beginning. However, it is also expected that certain parts will become clearer only when put in context with their application in the subsequent instructions section 6, or during the work in the lab course itself.

### 5.1 Computing and Python Comments

We are using the programming language Python in this lab course. Python is an easy to learn, powerful programming language. It has efficient high-level data structures and a simple but effective approach to object-oriented programming. Python's elegant syntax and dynamic typing, together with its interpreted nature, make it an ideal language for scripting and rapid application development in many areas on most platforms.

As already pointed out in section 1, basic programming skills in Python have been assumed and cannot be taught within the limited time available in this lab course. It is essential that you refresh or learn basic commands and coding elements in Python before starting this lab course. Links to two Python tutorials can be found in subsection 1.2.

The personal computers in this lab course are running Ubuntu 18.04 LTS as operating system. Ubuntu is a free and open-source Linux distribution based on Debian. As most modern operating systems, Ubuntu utilizes a graphical desktop environment, [GNOME](#), which is simple and easy to use such that sophisticated knowledge in Linux is not required.

In order to run a Python script, open a terminal by right-clicking onto the desktop and choosing “Open Terminal” from the drop-down menu. In the terminal window type:

```
$ python name_of_script.py
```

where `name_of_script` is the actual file name.

### 5.2 The Data Analysis Framework ROOT

ROOT started originally as a plotting program, which has matured into a multi-purpose data analysis framework used by many particle physicists. It can be operated in different ways: from the command line with the build-in C++ interpreter Cling, with interpreted C++ macros or C++ compiled code, or with Python scripts. For smaller data samples, the latter becomes more and more popular. Also in this lab course Python is used for both the data analysis as well as post-processing the results.

As already mentioned in subsection 1.3 on how to prepare for this lab course, extensive documentation is available on the [ROOT](#) homepage, including a primer, various tutorials, and a detailed user's guide. However, as also indicated previously, after some initial manual reading and tutorial exercises in order to get used to the basic ideas and syntax, most students learn ROOT by simply using it in the process of code development.

The largest problem with this “learning by going” approach is the difficulty to be aware of the wide range in functionality which ROOT has to offer, and to know about all the various classes and their abilities which are available in the ROOT libraries. When the task to solve is clear, a dedicated internet query using your favourite web search engine will usually deliver satisfying results on how to approach the problem in short amount of time.

The moment the name of the ROOT class is known to possibly fulfil the task you are interested in, the most useful web page is the [Class Index](#) which is part of the [ROOT Reference Guide](#). It can be found via the drop-down menu “All Classes”, which in general offers multiple ways to access and visualise the full class content of the ROOT framework. Note that the reference documentation is updated with each new ROOT release in order to incorporate latest changes.

The [Class Index](#) explains the use of each ROOT class, and also lists all member functions and their meanings. In the following a few examples are given for classes which undoubtedly will be used in this lab course. Please browse the links and make yourself familiar with the principle on how to use the content.

**TBrowser** Using a TBrowser one can browse all ROOT objects. It shows in a list on the left side of the window all browsable ROOT classes.

**TTree** A TTree object has a header with a name and a title. It consists of a list of independent branches ([TBranch](#)). Each branch has its own definition and list of buffers.

**TH1** Here all ROOT supported histogram types are explained (1-D, 2-D, 3-D, Profile). All histogram classes are derived from the base class TH1.

**THStack** The Histogram stack class. A THStack is a collection of TH1 or TH2 histograms. Using THStack::Draw() the histogram collection is drawn in one go according to the drawing option.

**TF1** A TF1 object is a 1-Dim function defined between a lower and upper limit. The function may be a simple function based on a [TFormula](#) expression or a precompiled user function.

**TLegend** This class displays a legend box ([TPaveText](#)) containing several legend entries. Each legend entry is made of a reference to a ROOT object, a text label and an option specifying which graphical attributes (marker/line/fill) should be displayed.

### 5.3 Monte Carlo Simulation

Monte Carlo (MC) simulation is an essential component of experimental particle physics in all the phases of its life-cycle: the investigation of the physics reach of detector concepts, the design of facilities and detectors, the development and optimization of data reconstruction software, the data analysis for the production of physics results [15].

The Monte Carlo simulation chain consists of three subsequent steps which are event generation, detector simulation and reconstruction. First, events are generated according to a chosen model. The so-called Monte Carlo event generator program calculates the cross-section of the specified physics process in a defined kinematic region. For the integration, a random number generator is used – hence the name Monte Carlo.

Usually the calculation can only be performed in leading order (LO) or next-to-leading order (NLO). Higher order processes are statistically simulated at parton level known as parton showering. When all particles have reached low energy, they are forced into (semi-stable) hadrons; this step is referred to as hadronisation. In the event generation process, all parent-daughter relations are kept track of, and the output is the complete four-vector information of all particles, also known as truth level. A sketch of a hadron-hadron collision as generated by a Monte Carlo program is depicted in figure 10.

In the simulation step, the full detector response is evaluated, including analogue signals, digitization and trigger. The simulation toolkit GEANT [17]–[19] uses the generator output (i.e. the particles’ four-vectors) and simulates the particles’ interactions with the detector volume. This includes particle

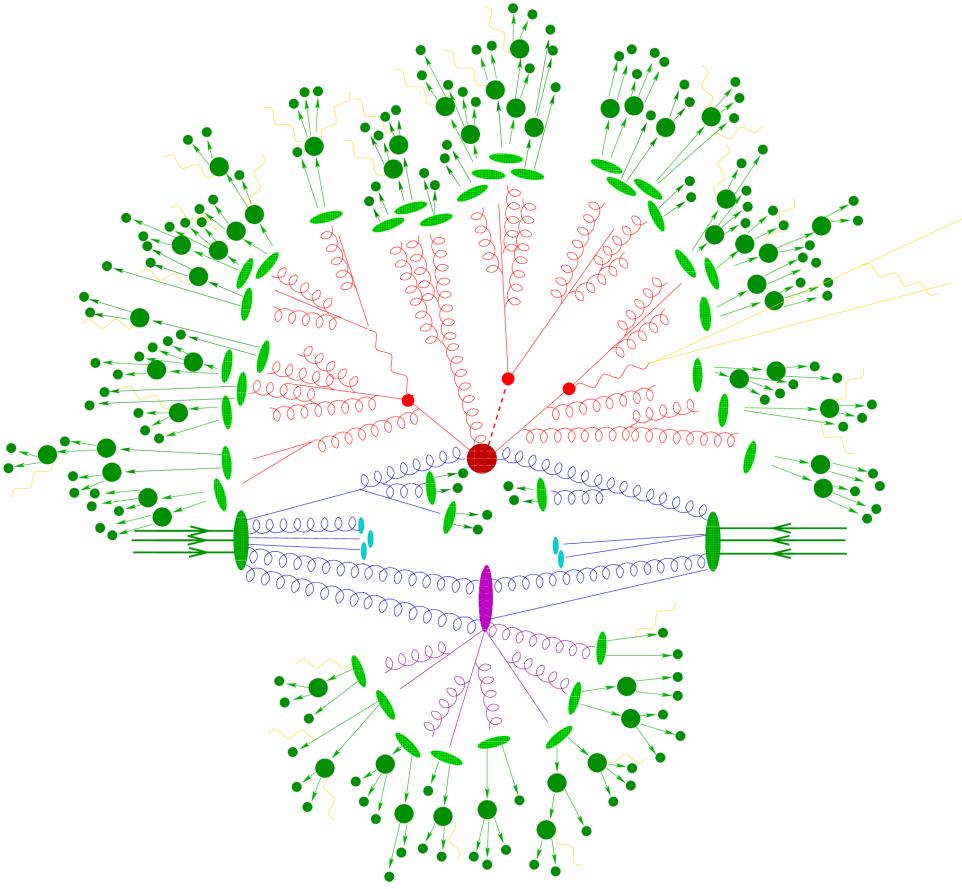


Figure 10: Sketch of a hadron-hadron collision as simulated by a Monte-Carlo event generator. The red blob in the center represents the hard collision, surrounded by a tree-like structure representing Bremsstrahlung as simulated by parton showers. The purple blob indicates a secondary hard scattering event. Parton-to-hadron transitions are represented by light green blobs, dark green blobs indicate hadron decays, while yellow lines signal soft photon radiation [16].

ionisation in trackers, energy deposits in calorimeters as well as intermediate particle decays and radiation. The output format of the simulation is identical to that of the real data and the same software is used to reconstruct particle candidates (electrons, photons, muons, etc.) in both cases, as described in subsections 4.2 and 4.3.

More details can be found in following reviews of the Particle Data Group:

<http://pdg.lbl.gov/2018/reviews/rpp2018-rev-mc-event-gen.pdf> ,  
<http://pdg.lbl.gov/2018/reviews/rpp2018-rev-monte-carlo-techniques.pdf> .

### Monte Carlo Particle Numbering Scheme

The Monte Carlo particle numbering scheme, also referred to as Particle Data Group Identifiers, PDGID, is intended to facilitate interfacing between event generators, detector simulators, and analysis packages used in particle physics. It is maintained by the Particle Data Group in cooperation with Monte Carlo authors.

The general form is a 7-digit number which encodes information about the particle's spin, flavour content, and internal quantum numbers. Particles are given positive numbers, antiparticles negative

numbers. Table 3 lists the PDGID for the six quarks and leptons of the Standard Model. They are numbered consecutively starting from 1 and 11, respectively. To do this they are first ordered by family and within families by weak isospin. More details can be found in this [PDG review](#).

Quarks						Leptons					
<i>d</i>	<i>u</i>	<i>s</i>	<i>c</i>	<i>b</i>	<i>t</i>	$e^-$	$\nu_e$	$\mu^-$	$\nu_\mu$	$\tau^-$	$\nu_\tau$
1	2	3	4	5	6	11	12	13	14	15	16

Table 3: The PDGID for the six quarks and leptons of the Standard Model.

## Event Weights and Scaling Factors

When a Monte Carlo sample is generated, the size of the sample is determined by requesting the generation of a certain number of events. The cross section of the sample is a fixed quantity dependent on the physics process. Since number of events, cross-section and integrated luminosity are related according to  $N = \sigma \cdot \mathcal{L}_{\text{int}}$  (equation 3) the corresponding luminosity of a MC sample varies according to the number of generated events, and the cross section of the underlying process.

As described in subsection 3.1, the amount of data collected in an experiment is expressed in terms of integrated luminosity since this quantity can be determined from known properties of the colliding beams. In order to perform a simulation of the data at given luminosity, it is necessary to scale the MC sample to correspond to this luminosity.

At a given centre-of-mass energy, the differential cross-section for a given physics process usually falls steeply as function of (transverse) energy. In order to populate the full kinematic phase-space in the simulation with sufficient number of events, the generation of Monte Carlo samples is hence often split into several parts, referred to as slices. The cut is placed on truth level variables, such as the generated transverse momentum  $p_T$  of the  $Z$  boson, which is required to exceed this cut value in the generation process of the corresponding slice. The full Monte Carlo sample is then assembled by combining all slices, properly taking into account the individual scaling factors.

In addition to performing the Monte Carlo generation in slices, this “weighting” of Monte Carlo events might follow a functional form to further flatten the response. In this case, the previously sample or slice dependent scaling factors become event level dependent weights, which are saved with each event in the generator output. These event weights need to be taken into account in addition to the overall luminosity scaling when comparing the distributions of the Monte Carlo simulation with those of real collision data.

## 5.4 Hadron Collider Event Kinematics

The underlying process in hadron-hadron collisions, either proton-proton or proton-antiproton collisions, is the interaction of two partons, which can be either quarks, antiquarks or gluons [1]. However, the momentum fractions  $x_1$  and  $x_2$  of the two interacting partons are unknown.

At a hadron collider, such as the LHC, the collisions take place in the centre-of-mass frame of the  $pp$  system, which is not the centre-of-mass frame of the colliding partons. The net longitudinal momentum of the colliding parton-parton system is given by  $(x_1 - x_2)E_p$  where  $E_p$  is the energy of the proton. Consequently, in a process such as  $pp \rightarrow 2 \text{jets} + X$ , the two final-state jets are boosted along the beam direction. For this reason, the jet angles are usually expressed in terms of rapidity  $y$ ,

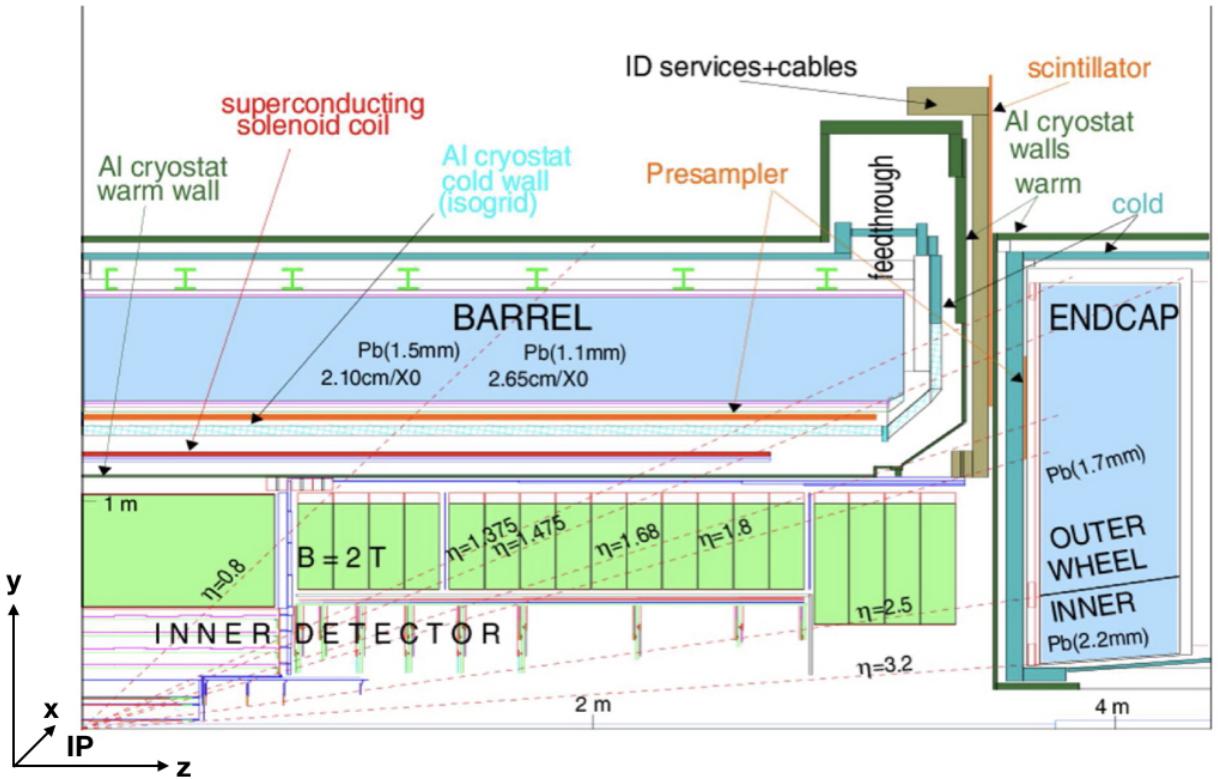


Figure 11: Longitudinal cut-out of the ATLAS Liquid Argon Calorimeter system in the  $y - z$  plane [20]. Shown is one quarter of the cross section through the calorimeter.

defined by

$$y = \frac{1}{2} \ln \frac{E + p_z}{E - p_z} ,$$

where  $E$  and  $p_z$  are the measured energy and z-component of momentum of a jet. The use of rapidity has the advantage that rapidity differences are invariant under boosts along the beam direction. Therefore the *a priori* unknown longitudinal boost of the parton-parton system does not affect the distribution of rapidity differences.

The invariant mass of the system of particles forming a jet is referred to as jet mass, which is not the same as the mass of the primary parton but mainly generated in the hadronisation process. For high-energy jets, the jet mass is usually small compared to the jet energy, and  $p_z \approx E \cos \theta$  where  $\theta$  is the polar angle of the jet with respect to the beam axis. When the jet mass can be neglected, the pseudorapidity defined as

$$\eta = -\ln \left( \tan \frac{\theta}{2} \right)$$

can be used in place of the rapidity  $y$ . Broadly speaking, the differential cross sections for jet production in hadron-hadron collisions are approximately constant in pseudorapidity, implying that roughly equal number of jets are observed in each interval of pseudorapidity.

Figure 11 shows the cross-section of the ATLAS Liquid Argon Calorimeter system through the beam axis, also illustrating the polar angle ranges covered by different regions of rapidity. The ATLAS detector design is approximately symmetrical in pseudorapidity. The detector region  $|\eta| > 3$  is particularly challenging since, for above explained reason, particle densities and energies are at their highest.

## 5.5 The Tag-and-Probe Method

The precise knowledge of the particle detection efficiencies plays a crucial role when for instance measuring cross sections or searching for new physics signatures (limit setting). Measuring the efficiencies requires a clean and unbiased sample of the particles under investigation. For electrons, the method of choice is the tag-and-probe method, which makes use of the characteristic signatures of  $Z \rightarrow ee$  decays.

In this method, strict selection criteria are applied on one of the two decay electrons, called tag, and the second electron, the probe, is used for the efficiency measurements. Standard event selection criteria are applied to further reject background. The tag-and-probe pairs must also pass requirements on their reconstructed invariant mass.

The probe samples might be contaminated by background objects (for example, hadrons misidentified as electrons, electrons from semileptonic heavy flavour decays or from photon conversions). This contamination can in principle be determined using either background template shapes or combined fits. However, the background contribution is reasonably small such that it can be neglected within the scope of this lab course.

The tag-and-probe selection makes sure, that the probe electron is most likely a real electron and thus can be used to determine detection efficiencies. For this the number of electrons is independently estimated at the probe level and at the level where the probe electron candidate satisfies in addition the tested criteria. The efficiency  $\epsilon$  is then defined as the fraction of probe electrons satisfying the tested criteria.

Usually the total efficiency to detect an electron in ATLAS is divided into different components, namely trigger, reconstruction and identification efficiencies, as well as the efficiency to satisfy additional analysis criteria, like isolation. The full efficiency  $\epsilon_{total}$  for a single electron can be written as:

$$\epsilon_{total} = \epsilon_{reconstruction} \times \epsilon_{identification} \times \epsilon_{trigger} \times \epsilon_{additional} .$$

## 5.6 ATLAS Open Data

ATLAS released an official dataset, open to the public for educational use only [3], [4]. The dataset consists of real data recorded with the ATLAS detector in 2012, with an integrated luminosity of  $(1.0007 \pm 0.0019) \text{ fb}^{-1}$  and a centre of mass energy of 8 TeV. Matching simulated Monte Carlo data for various physics processes have been made public too.

Both real and simulated data are subjected to a stringent preselection, both on event as well as on physics object level. In particular, a primary vertex is required, with more than four tracks pointing to it, and event cleaning as well as background rejection cuts are applied. A single lepton trigger must be satisfied, and the event must contain at least one reconstructed lepton with  $p_T > 25 \text{ GeV}$ .

On object level, medium++ identification quality is required for all electrons (c.f. subsection 4.2), and tight quality for all muons. In addition, electrons and muons must be well measured in the central part of the ATLAS detector with a minimum  $p_T$  of 5 GeV.

The released data is provided in a simplified format to reduce the complexity of a full scale analysis. The resulting format is a ROOT TTree with 45 branches as detailed in the appendix in table 4.

In addition to the datasets themselves, ATLAS released analysis tools as well as actual example analyses. Note however, that these are not needed for this lab course in which we will develop our own tools and analysis. Keeping this in mind, further general information can be found on the [public website](#) of the ATLAS Open Data project where also a [Get Started GitBook](#) is provided.

## 5.7 Uncertainties of the Measurement

In natural science, each measurement is performed with a certain precision, and the corresponding uncertainty or error must be indicated with the result. This is also – or in particular – true for experimental particle physics analyses in which it is not uncommon that a significant fraction of the effort is dedicated to the proper evaluation of the experimental errors.

One has to distinguish between statistical and systematic errors. Statistical errors are due to statistical uncertainties and arise from stochastic fluctuations. Important examples for measurements in experimental particle physics are finite statistics (Poisson distribution) and measurement resolutions.

Another class of errors, usually much harder to detect, are systematic errors which arise from uncertainties in the apparatus or model. Examples are calibration uncertainties, detector acceptance and poorly-known theoretical parameters. In general, systematic errors are the uncertainties in systematic correction factors.

“The treatment of systematic errors is one of the most demanding parts of experimental physics. It requires ingenuity, experience, cunning, and a fair degree of paranoia. These are well provided in particle physics. It also requires honesty and bravery, which is scarcer” (Roger Barlow).

## 6 Instructions

This section provides the actual instructions for the daily work during this lab course, which you should follow in the given order. After some general information about the files needed for the data analysis, we will start with a few exercises using ROOT interactively. While this is useful to get familiar with the structure of the underlying data – or more general for something “fast and simple” – the ROOT browser and command line are not suitable for systematically processing larger data sets. Therefore we’ll be moving on to using Python scripts rather quickly, in order to automate the work. When this is accomplished, we can dive deeper into the actual particle physics analysis.

### 6.1 Information about the Files in this Lab Course

All files which are needed for this lab course are stored in the directory `/fp/data`. These are the ATLAS data and simulated MC files in form of ROOT trees, as well as example Python code to analyse them. As in the real experiment, the data files are divided into so-called streams, depending on which triggers selected the event. For the MC files, the subdivision is based on the underlying physics process which was chosen for the event generation.

*Hint: The actual directory holding the files resides on a server and will be mounted automatically with AutoFS when accessed. Hence it is likely that the directory is not visible when you start your daily work. In this case, open a terminal (right-click with the mouse on the desktop and choose “Open Terminal” from the drop-down menu) and type `ls /fp/data`. This will cause the directory to be mounted, and also list its content.*

The ROOT files for the ATLAS data trees are stored in the directory `/fp/data/Data`. The `DataEgamma.root` file contains events selected by electron or photon triggers, while the `DataMuons.root` file contains events selected by muon triggers.

In the `/fp/data/MC` directory, various ROOT files for simulated MC processes are contained. Only three files are needed for this lab course: `mc_147770.Zee.root`, `mc_147771.Zmumu.root`, and `mc_147772.Ztautau.root`. As indicated by the names, these ROOT files contain events with  $Z$  bosons decaying into oppositely charged electron pairs, muon pairs, and tau pairs, respectively.

Finally, the folder `/fp/data/Code` contains example Python code for the analysis event loop, plotting as well as fitting the final results. These files should be copied into the user’s home area first day of the lab course, and can be modified there. Contrary, the ROOT files for the ATLAS data and MC simulations should remain in their original locations. Instead, the full path to the location of the files needs to be given to the Python analysis scripts when looping over the events.

### 6.2 Getting to Know the Basics

In this subsection you will be using ROOT interactively in order to make yourself familiar with the structure of the data files in the form of ROOT trees; and of course to get a first impression of ROOT itself. For this you will be using ROOT’s `TBrowser` which allows to browse ROOT objects in graphical way.

ROOT can be started from the command line, issuing the command `root` in a terminal window. The result is the output of the ROOT banner for a few seconds and the prompt of Cling which is an interactive C++ interpreter. User input is typed just after the ROOT prompt “`root[n]`” where `n` is counting the accepted commands. In order to quit this interactive ROOT session type “`.q`”.

1. Get familiar with the Open Data files provided for this course. In a terminal, open the DataEgamma.root file with ROOT and use the TBrowser to look at its content:

```
$ root /fp/data/Data/DataEgamma.root  
root [] TBrowser t
```

A short comment for each variable in the ROOT file can be found in Appendix 4 or online at [https://cheatham1.gitbooks.io/openatlasdatatools/variable\\_names.html](https://cheatham1.gitbooks.io/openatlasdatatools/variable_names.html).

2. Plot the vxp\_z variable. Make yourself clear what is shown, and why the distributions looks as it does. What does the statistics box in the upper right corner tell you? What are the units of the axis? Label them correctly by right-clicking with the mouse at the proper elements. Save and print the plot using the TBrowser: File → Save (or Save As). png or pdf are the file types of choice.
3. What are LHC bunches, and how are they related to the vxp\_z plot in question 2? How many peaks do you see in the vxp\_z distribution? Explain possible reasons. What type of plot could help you answering this question? *Hint: Maybe you can even find it under <https://twiki.cern.ch/twiki/bin/view/AtlasPublic/BeamSpotPublicResults>.*
4. Plot the lep\_n variable. What is shown? Unfortunately this distribution has too coarse binning out of the box. We need to use ROOT magic in the command line to show  $n = 1$  binning. Type  

```
root [] mini -> Draw("lep_n >> htemp(10,0,10)")
```

Then press enter and click with the middle mouse button on the canvas to refresh it.
5. How many leptons do you expect in the final state of a Z decay? Why do you see events here with more and less than two leptons? What is the dominant process for the 1-lepton final state? Draw Feynman diagrams for 1-lepton, 2-lepton and 3-lepton final states.
6. Plot the lep\_pt, lep\_eta and lep\_phi variables. What do you see? What does the statistics box in the upper right corner tell you now? Why are there more entries than in question 2? How many entries do you expect and why? What are the general features of the distributions? Why are there gaps in the  $\eta$  distribution? *Hint: Consult figure 11 and have a closer look on the tracker coverage as well as dead material distribution.* Where does the steep rise around 25 GeV in the  $p_T$  spectrum come from? Explain the concept of a trigger.

### 6.3 Automating Things

While interactive ROOT sessions are useful for quick checks and small operations, they are not suitable for systematic data analysis. Not only will the usually large size of the data sets become problematic, but more importantly, in data analysis we want to apply specific operations on certain observables (in the form of variables) for each event. Therefore a loop over all events in the ROOT tree is needed, allowing to program the desired operations at event level. For this we are using Python scripts in combination with PyROOT which is an extension module that allows the user to interact with any ROOT class from the Python interpreter.

1. Before you begin, create a work directory in your home area and copy all Python examples files from /fp/data/Code into that directory.
2. Look at the provided Python script eventloop.py and make yourself familiar with its content. What does it do? Run the script on the DataEgamma.root file and limit the number of events to a reasonable amount. Notice, that it writes an output file analysis\_xyz.root where xyz is

replaced by the name of your input file. It contains all histograms that have been defined in your `eventloop.py`. Practise the handling of primitive versus complex variable types such as arrays and lists by printing the content of various variables in the event loop, e.g. `vxp_z`, `lep_n`, `lep_pt`, `lep_eta`. Note that all energy and momentum variables are saved in MeV, but we want our final plots in GeV.

3. Compute the invariant mass of the two leading leptons in the event and plot it. Pay attention to the sign of  $M_{ll}^2$  and restrict to a physical result. *Hint: The leptons can be treated as massless. Recall the trigonometric relations between  $p_x$ ,  $p_y$ ,  $p_z$  and  $p_T$ . In particular show that  $p_z = p_T \cdot \sinh(\eta)$  by first evaluating its dependence on the polar angle  $\theta$ , and then rewriting as function of pseudorapidity  $\eta$ .*
4. Now use ROOT's `TLorentzVector` class to calculate the invariant mass and compare with your results by overlaying the two histograms. Make sure all plots have a legend and proper axis labels. *Hint: Use the histogram draw option "same" and change the plotting style of the curves either by using the ROOT editor, or directly in the Python script.*
5. How do you expect the distribution to look like for the decay of a  $Z$  boson – and why might it not look like your expectation when running over the `DataEgamma.root` file? How many peaks can you identify, and what do they belong to? *Hint: Try single and/or double logarithmic axis scales.* Now run your analysis over the Monte Carlo file `mc_147770.Zee.root` and compare the results for the invariant mass distribution.

## 6.4 Selecting Events

How can we improve the invariant mass distribution and extract a better  $Z$  boson peak? By applying a proper event selection! – In general one has to distinguish between cuts on event level, and selection criteria applied on particle level. Usually both are needed.

1. It is interesting to keep track of how many events are discarded by each criteria. This is called a “cut flow” histogram and very useful to have. Hence book a histogram with 20 bins that will show the total number of events remaining for each consecutive event selection criterion that will be introduced in the following.

ROOT allows to label the axis for each bin individually. So let's label each bin with the selection criterion it belongs to. Figure 12 shows as example the cut flow histogram for the analysis of the `DataEgamma.root` data file. Which are the two most effective cuts?

2. In the following, a brief description of the applied selection criteria is given. Implement them in your analysis code to select a clean sample of  $Z$  boson candidates. What do you think is the reason that the cuts on event level play only a minor role here? Validate the improvements of the cuts by monitoring the changes in the invariant mass distribution.

**Weights** This is not actually an event selection cut, but rather to keep track of the impact of event weights in the Monte Carlo files. While in real data a single event is unique and counts for one, event weights might have been applied in the MC generation (see subsection 5.3). Make sure to read out the `mcWeight` of each event and consider it in every filling of histograms. But note that this variable is not set in data and hence appears as 0. You need to take care that all data events are filled with a weight of 1 regardless.

**Trigger** In order to preselect events in which a  $Z$  boson decayed into an electron or muon pair, the corresponding trigger must be required to have fired the event. On this level also the luminosity accounting takes place.

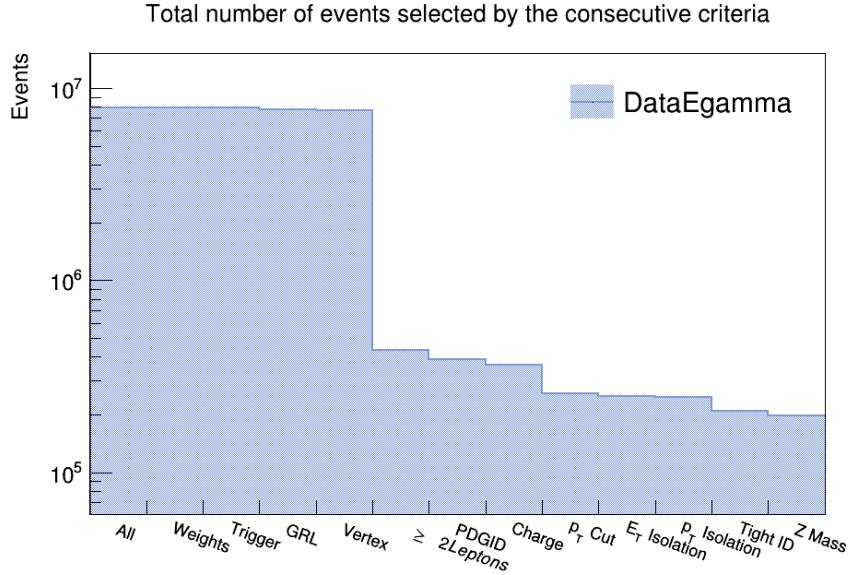


Figure 12: The cut flow for the selection of  $Z$  boson candidates for the `DataEgamma.root` file.

**GRL** The Good Run List. After the data acquisition of a so-called ATLAS run is completed, the data are investigated by the data quality group, and harsh criteria are applied in order to make sure that only meaningful data enter the physics analyses. This way whole runs, or in most cases parts of runs, in which for example important parts of the ATLAS detector were malfunctioning, will be flagged and can be excluded on analysis level (the subunit of an ATLAS run is called luminosity block and lasts approximately one minute).

**Vertex** In order to separate background from physics events, a well measured primary vertex is required.

**2 Leptons** Given that the  $Z$  boson decays into lepton pairs, at least two leptons are required.

**PDGID** Particle Data Group Identifiers. As described in subsection 5.3, each particle in Monte Carlo event generation is assigned to a unique identifier. Therefore also after the simulation of detector effects, the type of the particle is always known on MC level. For real data however, the identification of particles is more complicated and utilises the signatures as measured in the various detector components (c.f. section 4). To facilitate the particle identification, the Open Data project provides a generalised PDGID, `1ep_type`, which is valid both for MC as well as data. It can be used to uniquely identify the leptons as electrons and muons. *Hint: Antiparticles are typically assigned the same number as their corresponding particles, but with opposite sign. In your dataset, however, both particle and antiparticle receive the same positive number.*

**Charge** The  $Z$  boson as electrically neutral particle has to decay into a lepton pair with opposite charges. This can be taken into account in the event selection.

**$p_T$  Cut** Due to the high mass of the  $Z$  boson, the two decay leptons in the final state are expected to carry sizeable transverse momenta, with distributions peaking around half the  $Z$  mass. This can be utilised effectively in the event selection by requiring for the leptons  $p_T$  values of at least 25 GeV.<sup>14</sup>

<sup>14</sup> Also the thresholds for ATLAS' main single lepton triggers are placed such that they become fully efficient only in this  $p_T$  regime, making such transverse momentum cut advisable. Turning the argument around, the reason for this particular trigger setup is the aim to still be able to efficiently record the decays of the electroweak bosons.

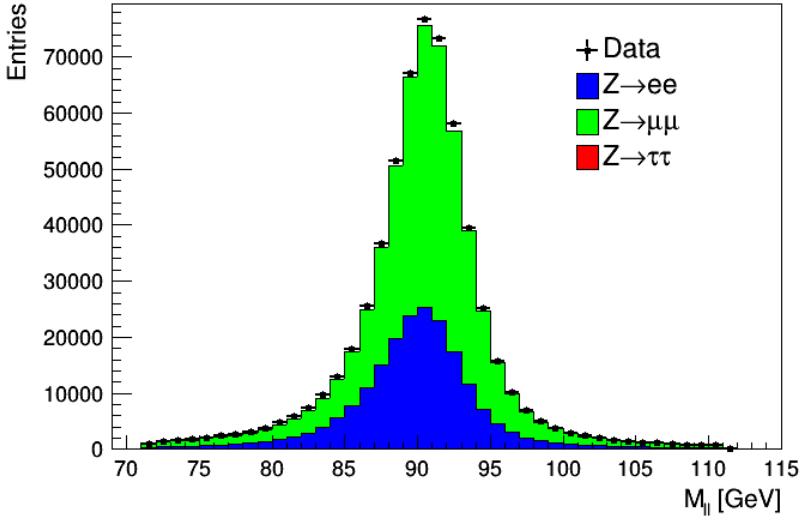


Figure 13: Invariant mass of the  $Z$  candidate.

As can be seen by monitoring the changes in the invariant mass distribution, the largest part of the background should be removed after having applied the lepton  $p_T$  cut (recall that the Open Data preselection applies rather harsh quality criteria on the leptons candidates, cf. subsection 5.6). However, as statistics is not of concern, it is still worth to increase the purity of the event sample by further tightening the lepton identification criteria. This can be achieved by introducing isolation cuts and increasing the lepton quality, as discussed in subsection 4.2.

**Isolation** The isolation in the  $\eta - \phi$  space can be used to improve the selection of leptons originating from  $Z$  boson decays. Calculate the *relative  $E_T$*  and  $p_T$  isolation of the leptons and investigate their distributions. *Hint: The amount of transverse energy (momentum) deposited in a cone around the lepton candidate should be compared to the transverse energy (momentum) carried by the lepton candidate itself.* What are reasonable cut values? Discuss your findings with the tutor.

**Tight ID** The lepton reconstruction algorithms on detector level provide the possibility to ask for looser or harsher identification criteria. These are bit-encoded in the `lept_flag` variable. Bit number 9 (when counting from zero) indicates a “tight” selection and should be required in the analysis. *Hint: Bit number 9 being set corresponds to an integer number  $2^9 = 512$ . You can check an arbitrary integer for its  $n$ 'th bit by performing a bitwise & with  $2^n$ .*

**Z Mass** Finally, the known mass of the  $Z$  boson can be used in the event selection, by restricting the invariant mass of the di-leptons to an allowed range.

3. Test the implementation of the event selection criteria continuously by processing a reasonably small sample of `DataEgamma.root` events. After completion, discuss your low-statistics invariant mass histogram with the tutor. If satisfactory, process the complete statistics of both data ROOT files and all three Monte Carlo files.

## 6.5 Comparing Data and MC Simulation

As described in subsection 5.3, Monte Carlo simulations play an important role in particle physics analysis. Amongst other they are used to prove that all detector effects are understood, and that all physics processes contributing to the final state under study are correctly modelled and included.

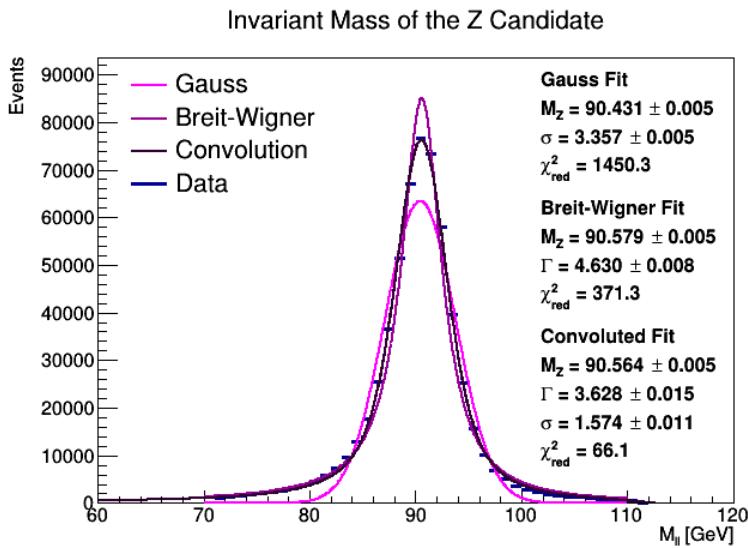


Figure 14: Fitting the invariant mass of the  $Z$  candidate.

A simple but descriptive way of demonstrating the level of agreement between data and simulation is the comparison of the measured and simulated distributions in “stacked plots”. Figure 13 shows such plot for the invariant mass of the  $Z$  candidate as measured in this lab course with the full statistics.

1. In figure 13 the data histogram corresponds to the sum of the  $Z$  boson candidates decaying into electron and muon pairs. To merge ROOT files containing histograms the utility `hadd` can be used. At the shell command line, simply type `hadd` to get online help.

```
$ hadd result.root file1.root file2.root ... filen.root
```

Use the `hadd` command to merge your histogram result files of the `Egamma` and `Muon` analyses into a combined Data results histogram file.

2. Compare your final invariant mass distribution of the merged data file to the Monte Carlo simulation. For this use and complete the provided script `plot.py` to make the same plot as shown in figure 13. With the `THStack` class, ROOT provides a simple way for plotting histograms stacked on top of each other. Note that proper scaling factors have to be applied to the Monte Carlo distributions to take into account the target data luminosity of  $1 \text{ fb}^{-1}$ , as described in subsection 5.3. Those are provided in the `plot.py` script.

## 6.6 Fitting the $Z$ Mass

The selected event sample can be used to study properties of the  $Z$  boson. Obviously, the invariant mass distribution provides a direct way for the determination of the  $Z$  boson mass and its decay width. For best results, the mass and width are extracted from a fit to the invariant mass spectrum, using the convolution of a Gauss and Breit-Wigner distribution, as shown in figure 14.

1. Why is the convolution of a Gauss and Breit-Wigner line shape expected to yield the best fit results? Which effects are taken into account by each of the two contributions?
2. Use and complete the provided script `fit.py` to extract the mass and width of the  $Z$  boson from the measured data sample. Compare your results with the [PDG](#) values.

## 6.7 Determining Efficiencies

The precise knowledge of efficiencies plays an important role in many areas of physics data analysis. Hereby not necessarily the question stands in the foreground, if or when a selection or measurement is fully efficient, but also how precise the efficiency is known. If errors are small, inefficiencies can be corrected for in the measurement. In this lab course, we determine the efficiency for the *tight* electron identification criteria.

1. Use the tag-and-probe method as described in subsection 5.5 to determine the detection efficiency for *tight* electrons in data as function of  $p_T$ . For the probe electron, disregard all other previously applied cuts in order to measure the efficiency of the *tight* identification. Fill a numerator and denominator histogram separately in the event loop. The final efficiency can be derived by dividing both histograms. *Hint: ROOT provides a TH1::Divide() method.*
2. Compare the *tight* electron detection efficiencies in data and Monte Carlo simulation. Is the detector simulation able to describe the data correctly?
3. What type of error determination needs to be applied in order to correctly account for the statistical uncertainties in the efficiency determination? The TH1::Divide function does not know a priori how the given histograms are related and which uncertainties to assign. But it accepts options to specify this.

## 6.8 Systematic Errors

So far there is no proper treatment of statistical and especially systematic errors implemented yet in this lab course – lucky you! However, considering the general comments in subsection 5.7 and recalling the individual analysis steps needed for the derivation of the  $Z$  boson mass – beginning with the calculation on event level up to the fitting of the invariant di-lepton mass distribution – you should be able to identify a variety of error candidates. Discuss with your tutor.

## Appendix

Branch Name	Type	Description
runNumber	int	Run identifier
eventNumber	int	Event identifier
channelNumber	int	Data sample ID eg WW sample 105985
mcWeight	float	Weight of a simulated event
pvp_n	int	Number of primary vertices
vxp_z	float	z-position of the primary vertex
trigE	bool	Boolean whether a standard trigger has fired in the egamma stream
trigM	bool	Boolean whether a standard trigger has fired in the muon stream
passGRL	bool	Signifies whether event passes the Good Run List may be put in isGoodEvent
hasGoodVertex	bool	Signifies whether the event has at least one good vertex where Ntracks > 4
lep_n	int	Number of preselected leptons
lep_truthMatched	vector<bool>	Boolean indicating whether the lepton is matched to a simulated lepton
lep_trigMatched	vector<bool>	Boolean signifying whether the lepton is the one triggering the event
lep_pt	vector<float>	Transverse momentum of the lepton
lep_eta	vector<float>	Pseudorapidity of the lepton
lep_phi	vector<float>	Azimuthal angle of the lepton
lep_E	vector<float>	Energy of the lepton
lep_z0	vector<float>	z-coordinate of the track associated to the lepton wrt. the primary vertex
lep_charge	vector<float>	Charge of the lepton
lep_type	vector<int>	Number signifying the lepton type (e, mu, tau) of the lepton
lep_flag	vector<int>	Bitmask implementing object cuts
lep_ptcone30	vector<float>	Scalar sum of track pTs in a cone of R=0.3 around lepton, not including lepton pT
lep_etcone20	vector<float>	Scalar sum of track ETs in a cone of R=0.2 around lepton, not including lepton ET itself
lep_trackd0pvunbiased	vector<float>	d0 of the track associated to the lepton at the point of closest approach (p.o.a.)
lep_tracksigd0pvunbiased	vector<float>	d0 significance of the track associated to the lepton at the p.o.a.
met_et	float	Transverse energy of the missing momentum vector
met_phi	float	Azimuthal angle of the missing momentum vector
jet_n	int	Number of selected jets
alljet_n	int	Total number of jets in event
jet_pt	vector<float>	Transverse momentum of the jet
jet_eta	vector<float>	Pseudorapidity of the jet
jet_phi	vector<float>	Azimuthal angle of the jet
jet_E	vector<float>	Energy of the jet
jet_m	vector<float>	Invariant mass of the jet
jet_jvf	vector<float>	Jet vertex fraction of the jet
jet_flag	vector<int>	Bitmask implementing object cuts of the top group
jet_trueflav	vector<int>	Flavor of the simulated jet
jet_truthMatched	vector<int>	Information whether the jet matches a simulated jet
jet_SV0	vector<float>	Weight from algorithm that reconstructs Secondary Vertices associated with a jet
jet_MV1	vector<float>	Weight from algorithm based on Multi-Variate technique
scaleFactor_PILEUP	float	Scalefactor for pileup reweighting. It effectively reweights the profile of average interactions per bunch crossing so that simulated data is the same as measured data.
scaleFactor_ELE	float	Scalefactor for electron efficiency
scaleFactor_MUON	float	Scalefactor for muon efficiency
scaleFactor_BTAG	float	Scalefactor for btagging algorithm. Should only be applied if analysis is specifically using b-tagging
scaleFactor_TRIGGER	float	Scalefactor to account for the different operating efficiencies of the used triggers
scaleFactor_JVFSF	float	Scalefactor for jet vertex fraction
scaleFactor_ZVERTEX	float	Scalefactor to reweight the distribution of the z position of the primary vertex

Table 4: Branches of the trees in the ATLAS Open Data ROOT files.

## References

- [1] M. Thomson, *Modern Particle Physics*. New York: Cambridge University Press, 2013, ISBN: 9781107034266. [Online]. Available: <http://www-spires.fnal.gov/spires/find/books/www?cl=QC793.2.T46::2013>.
- [2] ATLAS Collaboration, “The ATLAS Experiment at the CERN Large Hadron Collider,” *JINST*, vol. 3, S08003, 2008. DOI: [10.1088/1748-0221/3/08/S08003](https://doi.org/10.1088/1748-0221/3/08/S08003).
- [3] “Review Studies for the ATLAS Open Data Dataset,” CERN, Geneva, Tech. Rep. ATL-OREACH-PUB-2016-001, Aug. 2016. [Online]. Available: <https://cds.cern.ch/record/2203649>.
- [4] “Review of ATLAS Open Data 8 TeV datasets, tools and activities,” CERN, Geneva, Tech. Rep. ATL-OREACH-PUB-2018-001, Jun. 2018. [Online]. Available: <https://cds.cern.ch/record/2624572>.
- [5] M. Jende *et al.* (2018). International Masterclasses - Hands On Particle Physics, [Online]. Available: <http://atlas.physicsmasterclasses.org/en/index.htm> (visited on 12/21/2018).
- [6] P. M. Watkins, “Discovery of the W and Z Bosons,” *Contemp. Phys.*, vol. 27, pp. 291–324, 1986. DOI: [10.1080/00107518608211015](https://doi.org/10.1080/00107518608211015).
- [7] Richard Ruiz. (2015). Drell-Yan, Drell-Yan with Jets, Drell-Yan with all the Jets, [Online]. Available: <https://www.quantumdiaries.org/2015/05/18/dy-resummation/> (visited on 04/11/2019).
- [8] G. F. Giudice, “On Future High-Energy Colliders,” 2019. arXiv: [1902.07964 \[physics.hist-ph\]](https://arxiv.org/abs/1902.07964).
- [9] M. Wessels, “General Search for New Phenomena in  $ep$  Scattering at HERA,” PhD thesis, RWTH Aachen, Jul. 2004.
- [10] NobelPrize.org. Nobel Media AB. (2019). The Nobel Prize in Physics 1979, [Online]. Available: <https://www.nobelprize.org/prizes/physics/1979/summary/> (visited on 04/11/2019).
- [11] NobelPrize.org. Nobel Media AB. (2019). The Nobel Prize in Physics 1984, [Online]. Available: <https://www.nobelprize.org/prizes/physics/1984/summary/> (visited on 04/11/2019).
- [12] P. Zerwas, “W & z physics at lep,” *The European Physical Journal C - Particles and Fields*, vol. 34, no. 1, pp. 41–49, May 2004, ISSN: 1434-6052. DOI: [10.1140/epjc/s2004-01765-9](https://doi.org/10.1140/epjc/s2004-01765-9). [Online]. Available: <https://doi.org/10.1140/epjc/s2004-01765-9>.
- [13] CMS Collaboration. (2011). Invariant-mass spectra of opposite-sign muon pairs using 2010 and 2011 data, [Online]. Available: [https://twiki.cern.ch/twiki/bin/view/CMSPublic/PhysicsResultsMU0#Invariant\\_mass\\_spectra\\_of\\_opposi](https://twiki.cern.ch/twiki/bin/view/CMSPublic/PhysicsResultsMU0#Invariant_mass_spectra_of_opposi) (visited on 03/07/2019).
- [14] C. Lippmann, “Particle identification,” *Nucl. Instrum. Meth.*, vol. A666, pp. 148–172, 2012. DOI: [10.1016/j.nima.2011.03.009](https://doi.org/10.1016/j.nima.2011.03.009). arXiv: [1101.3276 \[hep-ex\]](https://arxiv.org/abs/1101.3276).
- [15] M. Grazia Pia and G. Weidenspointner, “Monte carlo simulation for particle detectors,” Jul. 2012. arXiv: [1208.0047 \[physics.comp-ph\]](https://arxiv.org/abs/1208.0047).
- [16] S. Höche, “Introduction to parton-shower event generators,” in *Proceedings, Theoretical Advanced Study Institute in Elementary Particle Physics: Journeys Through the Precision Frontier: Amplitudes for Colliders (TASI 2014): Boulder, Colorado, June 2-27, 2014*, 2015, pp. 235–295. DOI: [10.1142/9789814678766\\_0005](https://doi.org/10.1142/9789814678766_0005). arXiv: [1411.4085 \[hep-ph\]](https://arxiv.org/abs/1411.4085).
- [17] S. Agostinelli *et al.*, “Geant4—a simulation toolkit,” *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, vol. 506, no. 3, pp. 250–303, 2003, ISSN: 0168-9002. DOI: [https://doi.org/10.1016/S0168-9002\(03\)01368-8](https://doi.org/10.1016/S0168-9002(03)01368-8). [Online]. Available: <http://www.sciencedirect.com/science/article/pii/S0168900203013688>.

- [18] J. Allison *et al.*, “Geant4 developments and applications,” *IEEE Transactions on Nuclear Science*, vol. 53, no. 1, pp. 270–278, Feb. 2006, ISSN: 0018-9499. DOI: [10.1109/TNS.2006.869826](https://doi.org/10.1109/TNS.2006.869826).
- [19] J. Allison *et al.*, “Recent developments in Geant4,” *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, vol. 835, pp. 186–225, 2016, ISSN: 0168-9002. DOI: <https://doi.org/10.1016/j.nima.2016.06.125>. [Online]. Available: <http://www.sciencedirect.com/science/article/pii/S0168900216306957>.
- [20] M. Aharrouche *et al.*, “Study of the response of atlas electromagnetic liquid argon calorimeters to muons,” *Nucl. Instrum. Meth.*, vol. A606, pp. 419–431, 2009. DOI: [10.1016/j.nima.2009.05.021](https://doi.org/10.1016/j.nima.2009.05.021).